THE REVIEW OF SYMBOLIC LOGIC
Volume 9, Number 2, June 2016

EMBRACING THE TECHNICALITIES: EXPRESSIVE COMPLETENESS AND REVENGE

NICHOLAS TOURVILLE AND ROY T COOK

Abstract. The Revenge Problem threatens every approach to the semantic paradoxes that proceeds by introducing nonclassical semantic values. Given any such collection Δ of additional semantic values, one can construct a Revenge sentence:

This sentence is either false or has a value in Δ.

The *Embracing Revenge* view, developed independently by Roy T. Cook and Phlippe Schlenker, addresses this problem by suggesting that the class of nonclassical semantic values is indefinitely extensible, with each successive Revenge sentence introducing a new 'pathological' semantic value into the discourse. The view is explicitly motivated in terms of the idea that every notion that *seems* to be expressible (e.g., "has a value in Δ", for any definite collection of semantic values Δ) should, if at all possible, *be* expressible. Extant work on the Embracing Revenge view has failed to live up to this promise, since the formal languages developed within such work are expressively impoverished. We rectify this here by developing a much richer formal language, and semantics for that language, and we then prove an extremely powerful expressive completeness result for the system in question.

§1. Introduction: Revenge. We begin with the Liar paradox: the paradox that results from consideration of a sentence λ_1 that is equivalent to the assertion that λ_1 is false (i.e., in classical contexts, is not true):

$$\lambda_1 \leftrightarrow \neg T(\ulcorner \lambda_1 \urcorner),$$

where corner braces ($\ulcorner \ldots \urcorner$) represent Gödel coding or some other appropriate naming device. One extremely simple variant of the (classical, intuitionistic) derivation of a contradiction from the Liar sentence, plus the relevant instance of the *Tarskian T-schema*:

$$TS_{\Phi} : \Phi \leftrightarrow T(\ulcorner \Phi \urcorner)$$

goes as follows:

1	$\lambda_1 \leftrightarrow \neg T(\ulcorner \lambda_1 \urcorner)$	Liar sentence
2	$\lambda_1 \leftrightarrow T(\ulcorner \lambda_1 \urcorner)$	T-schema
3	$T(\ulcorner \lambda_1 \urcorner) \leftrightarrow \neg T(\ulcorner \lambda_1 \urcorner)$	1, 2, logic
4	$\bot$	3, logic.

We can express the same fact semantically as follows:[1]

PROPOSITION 1.1. *Given a classical base language $\mathcal{L}_0$ (i.e., a langauge $\mathcal{L}_0$ interpretable in terms of the classical truth values $\{\top, \bot\}$), extend $\mathcal{L}_0$ to $\mathcal{L}_1$ by adding a truth*

[1] A similar, albeit more complicated, result holds for intuitionistic languages and their semantics.

© Association for Symbolic Logic, 2016

doi:10.1017/S175502031600006X

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

predicate $T(x)$. Then λ_1, where λ_1 is equivalent to:

$$\neg T(\ulcorner \lambda_1 \urcorner)$$

cannot receive either of the values $\{\top, \bot\}$.

Much of the work on the semantic paradoxes focuses on 'solving' this conundrum (an informal version of *Tarski's Theorem*) where a solution can take one of (at least) four forms:

1. Reject one of the premises in the argument to the absurd conclusion.
2. Reject the logic used to obtain the absurd conclusion.
3. Reject one or more of the concepts involved in the argument to the absurd conclusion.
4. Accept the (presumably, on this approach, only apparently) absurd conclusion.[2]

Here we are interested in approaches to the Liar paradoxes, and to semantic paradoxes more generally, of the second type. In particular, we are interested in solutions that proceed by expanding the semantics via addition of a new 'pathological' truth value ρ_1.[3] The most famous such approach, of course, is the three-valued semantics developed by Saul Kripke in (Kripke 1975). On this account, sentences receive one of three semantic values $\top$, $\bot$, or ρ_1, the connectives obey the strong Kleene truth tables, and the truth predicate is assigned an extension and anti-extension via Kripke's well-known fixed-point construction.

Kripke's approach, although well-motivated and immensely important, suffers from an obvious defect (one he was explicitly aware of, see (Kripke 1975): 714 on "... the ghost of the Tarski hierarchy"): once the semantics has been extended via the addition of a new semantic value, the language is no longer expressively complete with respect to this new semantics. In particular, the language cannot be extended with a predicate (or sentential operator) that outputs the true when its argument takes the value ρ_1 and outputs the false otherwise:

PROPOSITION 1.2. *Extend the language $\mathcal{L}_1$ of Proposition 1.1 to $\mathcal{L}_2$ by adding a pathologicality predicate $P_1(x)$, or, alternatively, add a unary pathologicality operator $C_1(\cdots)$, so:*

$$P_1(x) =_{df} C_1(T(x)).$$

Then λ_2, where λ_2 is equivalent to:

$$\neg T(\ulcorner \lambda_2 \urcorner) \vee P_1(\ulcorner \lambda_2 \urcorner)$$

cannot receive any of the values $\{\top, \bot, \rho_1\}$.

This fact, or at least something equivalent to it, is more commonly formulated in terms of the impossibility of supplementing a language like $\mathcal{L}_1$ with a 'strong' negation "$\neg_S$"

[2] For an extended discussion of this four-part taxonomy of approaches to paradoxes, see (Cook 2013), especially Chapter 1.

[3] In this paper we shall, at least officially, be silent with regard to whether ρ_1 (or any of the ρ_αs discussed below) amounts to a genuine *semantic value*, on a par with truth and falsity, or whether we should, like Kripke, view ρ_1 (and all the ρ_αs discussed below) as merely a technically elegant device for representing the fact that sentences assigned this status in the formal semantics fail to receive a genuine semantic value. Unofficially, we prefer to interpret the infinite sequence of pathological values:

$$\rho_1, \rho_2, \rho_3, \ldots \rho_\alpha \ldots$$

as 'gaps' in Kripke's sense—that is, as (transfinitely many) different ways that a sentence can fail to receive a truth value, although the formal semantics sketched below is neutral with regard to such philosophical interpretation. For further discussion, see (Cook 2009).

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

(or, equivalently, with a 'weak' negation "$\neg_W$", which would allow for 'strong' negation as a defined notion)—that is, with either of the following:

Φ	$\neg_S\Phi$	$\neg_W\Phi$
$\top$	$\bot$	$\bot$
$\bot$	$\top$	$\top$
ρ_1	$\bot$	$\top$

Formulating the point in terms of a pathologicality predicate $P_1(x)$ (or pathologicality operator $C_1(\dots)$), however, is more useful for our purposes here.[4]

There is, of course, an obvious move to make at this point: If the addition of a truth predicate to $\mathcal{L}_0$ required the addition of an additional semantic value (ρ_1), then perhaps we can apply the same trick in order to interpret the language that has been further extended via the addition of a pathologicality predicate $P_1(x)$ by extending the semantics again, adding a second 'pathological' truth value ρ_2.

This 'trick' does indeed work—the addition of a second pathological value allows us to coherently assign semantic values to all sentences in $\mathcal{L}_2$ via a natural extension of Kripke's fixed point construction.[5] Unfortunately, the same 'revenge' phenomenon arises anew with regard to our new, expanded collection of four truth values:

PROPOSITION 1.3. *Let $\mathcal{L}_2$ be as in Proposition 1.2, and extend $\mathcal{L}_2$ to $\mathcal{L}_3$ by adding a pathological predicate $P_2(x)$, or, alternatively, add a unary pathologicality operator $C_2(\cdots)$, so:*

$$P_2(x) =_{df} C_2(T(x)).$$

Then λ_3, where λ_3 is equivalent to:

$$\neg T(\ulcorner\lambda_3\urcorner) \vee P_1(\ulcorner\lambda_3\urcorner) \vee P_2(\ulcorner\lambda_3\urcorner)$$

cannot receive any of the values $\{\top, \bot, \rho_1, \rho_2\}$.

Of course, we can address this problem as before, by adding a third pathological value ρ_3. But then the language $\mathcal{L}_3$ would be expressively incomplete with respect to this new semantics. Adding a new pathologicality predicate $P_3(x)$ in order to rectify this inadequacy would then require the addition of a fourth pathological value ρ_4. The cycle would continue.

The phenomenon just sketched is not limited to the finite. Even adding a countable infinity of new semantic statuses, or even any set-sized collection, will not eliminate the problem, as the following general theorem demonstrates:

THEOREM 1.4. *Given any definite collection of exclusive semantic values S, there is a sentence λ_S that cannot receive any of the values in S.*

Proof. For any definite collection of semantic values S, let $\Phi_S(x)$ be a predicate that holds of (the Gödel code, or other name of) a sentence Ψ if and only if the semantic value

[4] A pathologicality predicate $P_\alpha(x)$ and the corresponding pathologicality operator $C_\alpha(\dots)$ are not interchangeable or equivalent: given a pathologicality operator $C_\alpha(\dots)$, we can define a corresponding pathologicality predicate $P_\alpha(x)$ using the truth predicate as indicated above. We cannot in general define a pathologicality operator $C_\alpha(\dots)$ in terms of the relevant pathologicality predicate $P_\alpha(x)$, however, although we can mimic much of the logical behavior a pathologicality predicate via metalogical strategies such as Gödel coding (compare the obvious fact that we cannot, in general, define a falsity predicate in terms of negation).

[5] See (Cook 2008), (Cook 2009), (Schlenker 2010) for details.

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

of Ψ is in S (i.e., $\Phi_S(\ulcorner\Psi\urcorner)$ is $\top$ if the semantic value of Ψ is in S, and is $\bot$ otherwise). Then the required sentence is λ_S where λ_S is equivalent to:

$$\Phi_S(\ulcorner\lambda_S\urcorner) \wedge \neg T(\ulcorner\lambda_S\urcorner)).$$

Assume for *reductio* that the semantic value of λ_S is in S. Assume further that the value of λ_S is $\top$. Then the value of $\neg T(\ulcorner\lambda_S\urcorner))$ is $\bot$, so the value of $\Phi_S(\ulcorner\lambda_S\urcorner) \wedge \neg T(\ulcorner\lambda_S\urcorner)$ (i.e., the value of λ_S) is $\bot$. Contradiction. So the value of λ_S is in S, and is not $\top$. But then the value of $\Phi_S(\ulcorner\lambda_S\urcorner) \wedge \neg T(\ulcorner\lambda_S\urcorner)$ is $\top$, hence the value of λ_S is $\top$. Contradiction. So the value of λ_S cannot be in S. □

Assuming we do not abandon the addition-of-novel-semantic-status approach altogether, this super-strengthened version of the Revenge Phenomenon leaves us with two options:

1. Cut our losses, and accept that at some point our linguistic resources fall short of expressing all of these 'pathological' semantic notions (and, in particular, accept that there is some definite set of exclusive semantic values S such that $\Phi_S(x)$ is not expressible).
2. Accept the picture just sketched, and embrace revenge as an inherent feature of semantic theorizing itself, and of the language(s) within which we carry out such endeavors.

Most extant accounts of the semantics of truth (based on logical revision of some sort) amount to some form of option 1. For example, Graham Priest argues that certain forms of negation (including those that would be required to formulate predicates such as the various $\Phi_S(x)$s in Theorem 1.4 above) are not, in fact, coherent or expressible (see, e.g., (Priest 1990) for both an example of, and an insightful assessment of, this sort of approach). Here, however, we are interested in further exploring option 2. In particular, both the statement and the proof of Theorem 1.4 strike us (and, we suspect, many readers) as both coherent and, in fact, a legitimate piece of meta-logical reasoning. As a result, it would be preferable if we could formulate some account that allowed us to accept Theorem 1.4 (and the strong assumptions presupposed by Theorem 1.4 regarding what is legitimately expressible in both our informal languages and in the formal languages that model them), while also providing a satisfactory treatment of the Liar paradox and its closely related Revenge cousins. The basic building blocks of such an account—the *Embracing Revenge* view—have already been outlined by Roy T. Cook and Philippe Schlenker (independently) in (Cook 2008), (Cook 2009), and (Schlenker 2010).[6]

The Embracing Revenge account promises to provide an account that imposes no artificial limits on what can be said, while adequately dealing with the semantic paradoxes that *seem* to arise when no restrictions on expressivity are in place. In short, the Embracing Revenge view is motivated by the following rough *methodological maxim*:

> If Φ seems to be coherently expressible in our natural language(s), then, if at all possible, our formal accounts of those languages (and, in particular, our formal theories of truth and related semantic notions) should allow us to express Φ.

[6] Interestingly, (Cook 2008) and (Cook 2009) are primarily motivated by logical/philosophical concerns, while (Schlenker 2010) is motivated more by linguistic concerns (which we shall briefly discuss in §4 below). Thus it is particularly notable that these independent projects resulted in nearly identical accounts of the Revenge phenomenon.

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

Extant work on the Embracing Revenge view is explicitly motivated in terms of something like this maxim (see (Cook 2008), (Cook 2009), (Schlenker 2010)). But, as we shall see, the formalisms presented in this earlier work fall short of actually satisfying the goals informally codified in the maxim. Thus, the point of the remainder of this paper is to show how the Embracing Revenge view can deliver on this promise, and we shall accomplish this by:

1. Presenting a more expressive formalization of the Embracing Revenge view than given in previous work.
2. Proving a powerful expressive completeness result for this formalization.

Before doing so, however, a brief overview of the nuts-and-bolts of the Embracing Revenge View is in order—a task to which we now turn.

§2. Embracing Revenge: A Philosophical Sketch. We shall not attempt to rehearse all the important and interesting aspects of the Embracing Revenge view here, but shall instead merely outline the view and highlight those points that are directly relevant to the present project. The key thesis of the view is that we are, when semantically theorizing, continually expanding the expressive resources of our language along the lines sketched above (or, at the very least, we are when we are semantically theorizing *correctly*!) In particular, at any stage (after the first, where we supplement some purely classical language such as the language of standard Peano Arithmetic with a truth predicate) we are speaking a language $\mathcal{L}_\alpha$ that contains the standard logical vocabulary and also contains (some or all of) the semantic predicates:

$$\{T(x)\} \cup \{P_\beta(x) : 1 \leq \beta < \alpha\}.$$

Each such language $\mathcal{L}_\alpha$ can be coherently interpreted in terms of the set of semantic values:

$$\{\top, \bot\} \cup \{\rho_\beta : 1 \leq \beta \leq \alpha\}.$$

Each $\mathcal{L}_\alpha$ is obviously expressively impoverished, since it contains (at best) predicates for all semantic values *except* for the most 'recent' pathological value ρ_α. Thus, in order to express claims along the lines of:

Sentence Φ receives pathological value ρ_α,

where Φ is an expression of $\mathcal{L}_\alpha$, we need to 'semantically ascend' to a richer language, $\mathcal{L}_{\alpha+1}$, which does contain a predicate $P_\alpha(x)$, and hence can express such facts about the semantics of $\mathcal{L}_\alpha$. Of course, an adequate semantics for $\mathcal{L}_{\alpha+1}$ requires an additional pathological value $\rho_{\alpha+1}$, and as a result $\mathcal{L}_{\alpha+1}$ is unable to express some claims regarding its own semantics. Thus, in order to describe the semantics of $\mathcal{L}_{\alpha+1}$, we need to further ascend to $\mathcal{L}_{\alpha+2}$, and so on.

Fleshing this account out fully requires that we accept that:[7]

- The 'class' of semantic values:

$$\top, \bot, \rho_1, \rho_2, \cdots \rho_{\alpha-1}, \rho_\alpha, \rho_{\alpha+1}, \cdots$$

  is indefinitely extensible.

[7] The reader might wonder why we need a proper class of languages, of semantic values, and of semantic predicates, rather than merely a countably infinite (or, more generally, set-sized infinite) hierarchy (which seems to be all that the arguments given above explicitly motivate). The reason is that we can use the arithmetic (or set-theoretic, depending on the initial set-up) resources of the base language $\mathcal{L}_0$ in order to force 'large' jumps in the hierarchy. For example, if $\mathcal{L}_0$ is one's favorite standard classical first-order set theory, then in $\mathcal{L}_1$ we can formulate sentences along the lines of:

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

- The 'class' of semantic predicates:

$$T(x), P_1(x), P_2(x), \cdots P_{\alpha-1}(x), P_\alpha(x), P_{\alpha+1}(x), \cdots$$

  is indefinitely extensible.[8]
- The 'class' of languages:

$$\mathcal{L}_0, \mathcal{L}_1, \cdots \mathcal{L}_{\alpha-1}, \mathcal{L}_\alpha, \mathcal{L}_{\alpha+1}, \cdots$$

  is indefinitely extensible.

The notion of indefinite extensibility at work in the above clearly requires more attention than we will be able to give it here. Nevertheless, some quick comments are in order. The thought that many philosophically problematic concepts, including the concepts SET, ORDINAL NUMBER, and CARDINAL NUMBER, are indefinitely extensible traces back to Betrand Russell, who called such classes (and associated concepts) *self-reproductive*:

> ...the contradictions seem to result from the fact that...there are what we may call self-reproductive processes and classes. That is, there are some properties such that, given any class of terms all having such a property, we can always define a new term also having the property in question. Hence, we can never collect all of the terms having the said property into a whole; because, whenever we hope we have them all, the collection which we have immediately proceeds to generate a new term also having the said property. ((Russell 1906):144)

The term "indefinitely extensible", as well as the modern understanding of the notion, is due in large part to Michael Dummett's writings on the subject:

> The paradoxes—both the set theoretic and the semantic paradoxes—result from our possessing indefinitely extensible concepts...An indefinitely extensible concept is one for which, together with some determinate range or ranges of objects falling under it, we are given an intuitive principle whereby, if we have a sufficiently definite grasp of any one such range of objects, we can form, in terms of it, a conception of a more inclusive such range...By the nature of the case, we can form no clear conception of the extension of an indefinitely extensible concept; any attempt to do so is liable to lead us into contradiction. ((Dummett 1993): 454)

Interestingly, Dummett argued that it was not only set- or class-theoretic notions such as SET, ORDINAL, and CARDINAL that were indefinitely extensible, but that all paradox-prone notions, including truth, involved indefinite extensibility in some sense. Although the

---

> This sentence has one of the truth values that can be modelled by my current set theory other than the True.

This sentence must receive a truth value that cannot be described in that first-order set theory - i.e., ρ_α where α is the least ordinal not in the smallest standard model of $\mathcal{L}_0$. For further discussion see (Cook 2008).

[8] Equivalently, if the truth predicate is present (and we shall assume that it is throughout), the 'class' of pathologicality operators:

$$C_1(\cdots), C_2(\cdots), \cdots C_{\alpha-1}(\cdots), C_\alpha(\cdots), C_{\alpha+1}(\cdots), \cdots$$

is indefinitely extensible.

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

details of Dummett's analysis and treatment of such notions (in short: adopt intuitionistic logic when reasoning about indefinitely extensible concepts) is less than satisfactory when applied to the semantic paradoxes (see, e.g., (Williamson 1998) for discussion), we take the Embracing Revenge view to be in the spirit of Dummett's approach, while differing substantially in the details.

We can flesh out the notion of indefinite extensibility as follows:

DEFINITION 2.1. *A concept C is* indefinitely extensible *if and only if there is a function or operation f mapping collections of objects to objects such that, given any definite collection Δ of objects falling under C:*

$$f(\Delta) \in C$$

and:

$$f(\Delta) \notin \Delta.$$

With this definition in mind, we can now understand Theorem 1.4. above as nothing more than a demonstration that the concept SEMANTIC VALUE is indefinitely extensible: Given any definite collection (i.e., set) of semantic values S, the proof, in effect, provides a construction that generates a new semantic value not in S – namely the semantic value of the sentence λ_S where λ_S is equivalent to:

$$\Phi_S(\ulcorner \lambda_S \urcorner) \wedge \neg T(\ulcorner \lambda_S \urcorner)$$

(with $\Phi_S(x)$ defined as before). In addition, if, for every language $\mathcal{L}_\alpha$ with semantic vocabulary:

$$\{T(x)\} \cup \{P_\beta(x) : 1 \leq \beta < \alpha\}$$

(or equivalent) and semantic values:

$$\{\top, \bot\} \cup \{\rho_\beta : 1 \leq \beta \leq \alpha\}$$

there is a second language $\mathcal{L}_{\alpha+1}$ with semantic vocabulary:

$$\{T(x)\} \cup \{P_\beta(x) : 1 \leq \beta < \alpha + 1\}$$

(or equivalent) and semantic values:

$$\{\top, \bot\} \cup \{\rho_\beta : 1 \leq \beta \leq \alpha + 1\}$$

then we obtain as a corollary that the sequences of languages, and the sequence of semantic predicates that we piecewise add to obtain these languages, are themselves indefinitely extensible.[9]

Of course, to be a viable account of semantic notions—one not susceptible to Revenge-type worries—it had better not be the case that the view allows one to construct a 'super' Revenge sentence along the lines of:[10]

> This sentence has one of the truth values (i.e., one of *all of them*) other than the true.

[9] Actually, in order to secure the indefinite extensibility (rather than merely the infinity) of the sequence of languages $\mathcal{L}_\alpha$, we also need it to be the case that, given any (definite—i.e., set-sized) well-ordering of such languages, there is a further language that contains, or contains resources equivalent to, their union.

[10] Schlenker, in his eponymous paper (Schlenker 2010) calls such sentences *Super Liars*.

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

On the informal picture just sketched (and in the formal semantics developed below) this sentence is not expressible. While it is the case that, for any semantic value, there is a language $\mathcal{L}_\alpha$ 'high' enough in the hierarchy to 'talk' about that semantic value, there is no language 'high' enough that one can talk about *all* semantic values. In short, there is no 'last' language.

This is, of course, as it should be: the Embracing Revenge view is motivated by the thought that we are always in a language that cannot fully describe its own semantics, but we can always 'ascend' to a new, expressively richer language that can describe the semantics of the former language. But it is worth noting that the view commits us to rejecting absolutely general quantification—that is, we can never say "all semantic values" or "all languages", etc. and actually quantify over *all* of them. At first glance, this seems to conflict with the methodological maxim put forth in the previous section, but only apparently. We take that maxim to mean that we should, whenever possible, avoid rejecting certain notions as incoherent or inexpressible *solely* because they seem to generate Revenge-style paradoxes (or other sorts of pathology). If there are independent reasons for doubting that a particular notion is, in fact, genuinely expressible (regardless of the pre-theoretic utility or intuitiveness of the notion) then we have every reason to reject that notion. In fact, we believe there *are* independent reasons for rejecting absolutely general quantification over indefinitely extensible concepts generally (in fact, it is tempting to think that indefinitely extensible concepts can be alternatively defined as exactly those concepts over which we cannot unrestrictedly quantify).[11] As a result, the Embracing Revenge view involves no restrictions on what can be expressed other than those imposed by the indefinite extensibility underpinning the view itself.

Given that one of the central advantages, and important attractions, of the Embracing Revenge view is the claim that there is nothing (coherent) which we cannot say in any language, we need some reassurance that formal models of the approach in question (as well as the philosophical account of natural language semantics built upon such formalisms) do indeed allow us to express anything that can be expressed (that is, the models should allow us to express anything that is not ruled out as incoherent based on indefinite extensibility considerations). In short, we need some sort of expressive completeness result.

Before moving on to the formal model, and the proof of such an expressive completeness result for that formalism, it is worth making a few observations. First, despite the programmatic claims made in (Cook 2008), (Cook 2009), and (Schlenker 2010), the formal languages developed or discussed in those works are inadequate for the task at hand. The reason is simple. Although much of the formal work within (Cook 2008) and (Cook 2009) is similar in spirit to that found in (Schlenker 2010), the two parallel developments of the approach utilize different sets of connectives (Schlenker, for example, uses a generalization of strong Kleene conjunction and disjunction, while Cook uses a generalization of

[11] For a number of good discussions of the connections between indefinite extensibility and absolutely general quantification, see the essays in (Rayo & Uzquiano 2007). Rayo's own contribution to the volume (Rayo 2007) is particularly relevant here, as he argues that the difficulties with formulating a model theory for a particular language $\mathcal{L}$ within $\mathcal{L}$ itself presents a dilemma: either we pessimistically abandon the thought that every legitimate language can be given an adequate model theory, or we accept that there is an indefinitely extensible hierarchy of expressively richer languages. Rayo tentatively suggests that the latter option is perhaps the correct one, in broad agreement with the general approach of this paper, although there is no indication that Rayo takes these languages to be anything other than fully classical (see pp. 248–256). For more on how these issues play out in the Embracing Revenge account see (Cook 2008).

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

weak Kleene conjunction and disjunction). Further, Schlenker proves that the connectives found in (Cook 2008) are not definable in terms of his connectives, and vice versa (see (Schlenker 2010): 410). Thus, any attempt to prove an appropriate expressive completeness result for either of the extant languages developed under the Embracing Revenge banner is doomed to fail, since they each fail to express notions that can be expressed on the view (and are expressed, in the competing account).

Thus, in order to prove an appropriate version of the expressive completeness result, and thus defend the claim that the Embracing Revenge view allows us to say everything that can be said, we need first to formulate a (sequence of) formal language(s) that is significantly richer, expressively, than those found in (Cook 2008), (Cook 2009), and (Schlenker 2010). Developing such a framework is the task of the next section.

§3. Embracing Revenge: The Formalism. We begin with a well-ordered sequence of languages[12] $\mathcal{L}_\alpha$ (one for each ordinal $\alpha \in ON$):[13]

$$\mathcal{L}_0 \subseteq \mathcal{L}_1, \subseteq \cdots \subseteq \mathcal{L}_\alpha \subseteq \mathcal{L}_{\alpha+1}, \cdots$$

The primitive expressions of each $\mathcal{L}_\alpha$ are as follows:

- $\aleph_\alpha$-many propositional letters P_β
  (where β is an ordinal such that $0 \leq \beta < \aleph_\alpha$)
- $\aleph_\alpha$-many sentence names S_β
  (where β is an ordinal such that $0 \leq \beta < \aleph_\alpha$)
- A sequence of propositional constants C_β, $0 \leq \beta \leq \alpha$
- Negation ($\neg$)
- A sequence of 'relativized' conditionals $\rightarrow_\beta$, $0 < \beta \leq \alpha$[14]
- $\aleph_\alpha$-ary (infinitary) conjunction ($\wedge$)
- A truth predicate $T(x)$ (except for $\mathcal{L}_0$)

The formation rules are as one would expect (but note the cardinality restriction on conjunction):

- If $0 \leq \beta < \aleph_\alpha$, then P_β is a wff of $\mathcal{L}_\alpha$.
- If $0 \leq \beta < \aleph_\alpha$, then $T(S_\beta)$ is a wff of $\mathcal{L}_\alpha$.
- If $0 \leq \beta \leq \alpha$, then C_β is a wff of $\mathcal{L}_\alpha$.

[12] Philosophically, the view is best understood as involving a single, but constantly expanding language. From a technical perspective, however, it is more convenient to construct a series of distinct languages, where $\mathcal{L}_\beta \subseteq \mathcal{L}_\alpha$ if $\beta \leq \alpha$.

[13] We use ON as an informal abbreviation for the class of all ordinals. Note that there is a systematic ambiguity here: Informally, the view is intended to describe a sequence of languages, and corresponding sequence of semantic predicates and semantic values, of the same 'length' as the collection of *all* ordinals. But the impossibility of quantifying over the entirety of an indefinitely extensible collection prevents our quantifying over all ordinals as well. Hence, in the formal development of the Embracing Revenge account, quantification over 'all' ordinals is always relativized to a particular background set theory being mobilized in the metatheory—that is, we are always quantifying over those ordinals whose existence is entailed by the set theory currently being mobilized, and we are thus always describing a partial and incomplete (albeit extendable via stronger set theoretical assumptions) formal model of the Embracing Revenge account.

[14] Note that $\mathcal{L}_0$ does not contain a primitive conditional, since the 'classical' material conditional for $\mathcal{L}_0$ is definable as:

$$\Phi \rightarrow_m \Psi = \neg(\Phi \wedge \neg\Psi).$$

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

- If Φ is a wff of $\mathcal{L}_\alpha$, then $\neg\Phi$ is a wff of $\mathcal{L}_\alpha$.
- If Φ and Ψ are wffs of $\mathcal{L}_\alpha$, and $0 < \beta \leq \alpha$, then $\Phi \rightarrow_\beta \Psi$ is a wff of $\mathcal{L}_\alpha$.
- If Σ is a set of wffs of $\mathcal{L}_\alpha$ and $|\Sigma| \leq \aleph_\alpha$, then $\wedge(\Sigma)$ is a wff of $\mathcal{L}_\alpha$.

It is worth pointing out that $\mathcal{L}_\alpha$ does not contain quantifiers (solely for technical convenience). Given any particular (definite) domain, however, universal quantification can be 'mimicked' by replacing the propositional letters P_β with primitive term/ primitive predicate pairs, and then using infinitary conjunctions and disjunctions of sufficiently generous arity (i.e., in some $\mathcal{L}_\beta$ where β is sufficiently large) as 'substitutional' analogues of universally and existentially quantified formulas respectively. If we were to extend the language with the explicit addition of a primitive quantifier, then the universal quantifier would be handled analogously to the infinitary conjunction, with the existential quantifier defined in terms of the universal quantifier and negation (or vice versa—see the definition of disjunction below).[15]

Before assigning semantic values to sentences, we need to know what our sentence names denote. Hence, any model of an $\mathcal{L}_\alpha$ begins with a denotation function:

DEFINITION 3.1. *A* denotation function $\delta_{\mathcal{L}_\alpha}$ *is any function mapping the sentence names of* $\mathcal{L}_\alpha$ *to the wffs of* $\mathcal{L}_\alpha$.

The assignment of semantic values to sentences of $\mathcal{L}_\alpha$ (relative to $\delta_{\mathcal{L}_\alpha}$) begins with a truth-value assignment on sentence names and propositional letters:

DEFINITION 3.2. *An* assignment $\sigma_{\mathcal{L}_\alpha}$ *on* $\mathcal{L}_\alpha$ *is any function from the sentence names and propositional letters of* $\mathcal{L}_\alpha$*:*

$$\{S_\beta : 0 \leq \beta < \aleph_\alpha\} \cup \{P_\beta : 0 \leq \beta < \aleph_\alpha\}$$

to the set of truth values:

$$\{\top, \bot\} \cup \{\rho_\beta : 0 < \beta \leq \alpha\}.$$

We extend assignments on the sentences names of $\mathcal{L}_\alpha$ to valuations on the entire language as follows:

DEFINITION 3.3. *A* valuation $\sigma^*_{\mathcal{L}_\alpha}$ *on* $\mathcal{L}_\alpha$ *is obtained from an assignment* $\sigma_{\mathcal{L}_\alpha}$ *on* $\mathcal{L}_\alpha$ *via the following rules:*

$$\sigma^*_{\mathcal{L}_\alpha}(P_\beta) = \sigma_{\mathcal{L}_\alpha}(P_\beta).$$

$$\sigma^*_{\mathcal{L}_\alpha}(T(S_\beta)) = \sigma_{\mathcal{L}_\alpha}(S_\beta).$$

$$\sigma^*_{\mathcal{L}_\alpha}(C_0) = \bot$$
$$\sigma^*_{\mathcal{L}_\alpha}(C_\beta) = \rho_\beta \text{ if } \beta > 0.$$

$$\begin{aligned} \sigma^*_{\mathcal{L}_\alpha}(\neg\Phi) = \top \quad & \text{If } \sigma^*_{\mathcal{L}_\alpha}(\Phi) = \bot \\ = \bot \quad & \text{If } \sigma^*_{\mathcal{L}_\alpha}(\Phi) = \top \\ = \sigma^*_{\mathcal{L}_\alpha}(\Phi) \quad & \text{Otherwise.} \end{aligned}$$

[15] For a somewhat more detailed treatment of the addition of quantifiers to the formal system, see (Schlenker 2010): 400–402.

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

$$\begin{aligned}
\sigma^*_{\mathcal{L}_\alpha}(\wedge(\Sigma)) \quad &= \top \ \textit{If } \sigma^*_{\mathcal{L}_\alpha}(\Phi) = \top \textit{ for all } \Phi \in \Sigma. \\
&= \bot \ \textit{If } \sigma^*_{\mathcal{L}_\alpha}(\Phi) = \bot \textit{ for some } \Phi \in \Sigma. \\
&= \rho_\beta \ \textit{Otherwise,} \\
&\qquad \textit{where } \beta = sup\{\gamma : \Phi \in \Sigma \textit{ and } \sigma^*_{\mathcal{L}_\alpha}(\Phi) = \rho_\gamma\}.
\end{aligned}$$

$$\begin{aligned}
\sigma^*_{\mathcal{L}_\alpha}(\Phi \rightarrow_\beta \Psi) \quad &= \top \ \textit{If } \sigma^*_{\mathcal{L}_\alpha}(\Phi) \in \{\top, \bot\} \cup \{\rho_\beta : \beta < \alpha\} \\
&\qquad \textit{and } \sigma^*_{\mathcal{L}_\alpha}(\Phi) \leq \sigma^*_{\mathcal{L}_\alpha}(\Psi) \textit{ relative to the ordering:} \\
&\qquad \cdots \rho_{\gamma+1} < \rho_\gamma < \rho_{\gamma-1} < \cdots < \rho_2 < \rho_1 < \bot < \top. \\
&= \bot \ \textit{If } \sigma^*_{\mathcal{L}_\alpha}(\Phi) \in \{\top, \bot\} \cup \{\rho_\beta : \beta < \alpha\} \\
&\qquad \textit{and } \sigma^*_{\mathcal{L}_\alpha}(\Phi) > \sigma^*_{\mathcal{L}_\alpha}(\Psi) \textit{ relative to the ordering:} \\
&\qquad \cdots \rho_{\gamma+1} < \rho_\gamma < \rho_{\gamma-1} < \cdots < \rho_2 < \rho_1 < \bot < \top. \\
&= \rho_\zeta \ \textit{Otherwise,} \\
&\qquad \textit{where } \sigma^*_{\mathcal{L}_\alpha}(\Phi) = \rho_\eta, \sigma^*_{\mathcal{L}_\alpha}(\Psi) = \rho_\theta, \textit{ and } \zeta = max\{\eta, \theta\} \\
&\qquad \textit{relative to the ordering:} \\
&\qquad \cdots \rho_{\gamma+1} < \rho_\gamma < \rho_{\gamma-1} < \cdots < \rho_2 < \rho_1 < \bot < \top.
\end{aligned}$$

Note that valuations, unlike assignments, only have well-formed formulas in their range. The valuation corresponding to a particular assignment is independent of the denotation function (the denotation function *is* relevant to determining if that assignment is *acceptable*—see below). Finally, infinitary conjunctions may receive a value different from the values received by any conjunct, since in the case covered by the third clause the conjunction takes, as value, the *supremum* (loosely speaking) of the ρ_αs assigned to the conjuncts. For example, if Φ is a countably infinite conjunction:

$$P_1 \wedge P_2 \wedge P_3 \cdots \wedge P_{n-1} \wedge P_n \wedge P_{n+1} \wedge \ldots$$

where, for each $n \in \omega$, $\sigma_{\mathcal{L}_\omega}(P_n) = \rho_n$, then $\sigma^*_{\mathcal{L}_\omega}(\Phi) = \rho_\omega$.

An 'historical' note: This particular list of logical vocabulary, and the semantic clauses that go with that vocabulary, was arrived at as follows: The use of sentence names, propositional letters, and the treatment of the truth predicate as transparent is common to both (Cook 2008) and (Schlenker 2010). The conjunction given above is an infinitary generalization of the conjunction used in (Schlenker 2010). Negation is taken verbatim from (Cook 2008).[16] The definition of the conditional is novel to the present approach, as is the addition of pathologically constants.

Finally, we define the acceptability of an assignment (and corresponding valuation) relative to a particular denotation function as follows:

DEFINITION 3.4. *An assignment $\sigma_{\mathcal{L}_\alpha}$ (with corresponding valuation $\sigma^*_{\mathcal{L}_\alpha}$) on $\mathcal{L}_\alpha$ is* acceptable *on $\delta_{\mathcal{L}_\alpha}$ if and only if, for any S_β ($0 \leq \beta < \alpha$):*

$$\sigma_{\mathcal{L}_\alpha}(S_\beta) = \sigma^*_{\mathcal{L}_\alpha}(\delta_{\mathcal{L}_\alpha}(S_\beta)).$$

[16] (Schlenker 2010) does consider an alternative but equivalent formulation of his formal theory that uses the negation given above, plus the relativized strict truth predicates (see §5 below), as primitive.

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

In short, an assignment (and corresponding valuation) is acceptable if and only if it assigns the same semantic value to a sentence name and the formula that it denotes. A model of $\mathcal{L}_\alpha$ is then just a pair consisting of a denotation function and acceptable assignment:

DEFINITION 3.5. $\mathcal{M} = \langle \delta_{\mathcal{L}_\alpha}, \sigma_{\mathcal{L}_\alpha} \rangle$ *is a* model *of* $\mathcal{L}_\alpha$ *if and only if* $\sigma_{\mathcal{L}_\alpha}$ *is acceptable on* $\delta_{\mathcal{L}_\alpha}$.

There are two important things to note at this stage. First, it is important that all connectives are defined not only for the values that are in play in the semantics for the language in which that connective is first introduced, but that they are defined for all possible semantic values, since notions (such as the truth predicate $T(x)$ and the relativized conditionals $\rightarrow_\alpha$) that are introduced in one language continue to be present in 'later' extensions of the language, and thus must be coherently applicable in such extended contexts where 'new' semantic values are in play.[17] Second, note that C_0 is just a falsity constant, and each C_α where $\alpha > 0$ is a ρ_α-constant.[18] We will often write C_0 as $\bot$.[19]

Glancing at partial truth tables for these connectives (or analogues of truth tables, for semantic predicates) will help to make their properties more apparent. First, a sentence consisting of the application of the truth predicate to a sentence name receives the same truth value (according to the valuation σ^*) as does the sentence name itself (according to the corresponding valuation σ)—in short, the truth predicate is *transparent*:[20]

S_α	$T(S_\alpha)$
$\top$	$\top$
$\bot$	$\bot$
ρ_1	ρ_1
ρ_2	ρ_2
$\vdots$	$\vdots$
ρ_β	ρ_β
$\vdots$	$\vdots$

The primitive negation operator maps $\top$ to $\bot$, maps $\bot$ to $\top$, and maps any ρ_α to itself (we shall see other, defined, negations in §5):

[17] This fact also highlights one of the advantages of the present approach over the superficially similar hierarchy-based approach found in (Tarski 1933): On the present view semantic predicates such as the truth predicate meaningfully apply to all sentences, and not just to sentences 'lower' in the hierarchy.

[18] The fact that C_α is present in $\mathcal{L}_\alpha$ shows that it is not the case that we cannot say *anything* about ρ_α in $\mathcal{L}_\alpha$. What is prohibited is the presence, within $\mathcal{L}_\alpha$, of a *predicate* holding of exactly the sentences in $\mathcal{L}_\alpha$ that receive ρ_α as their semantic value.

[19] Note that, as a result, $\bot$ is (systematically and unproblematically) ambiguous, since it denotes both a (defined) expression in the object language and a semantic value (i.e., falsity) in the metatheory. It is always clear from context what is meant, however.

[20] As noted earlier, this is not, strictly speaking, a truth table, since it codifies the outcome of applying a predicate to a sentence name, rather than the result of applying a connective to a sentence. Similar comments apply to the tables given for other defined predicates discussed below, including the transparent falsity predate, the relativized strict truth and strict falsity predicates, and the pathologically predicates.

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

Φ	$\neg\Phi$
$\top$	$\bot$
$\bot$	$\top$
ρ_1	ρ_1
ρ_2	ρ_2
$\vdots$	$\vdots$
ρ_β	ρ_β
$\vdots$	$\vdots$

The binary version of conjunction is a generalization of the strong Kleene conjunction:[21] it outputs $\top$ if both conjuncts receive $\top$, outputs $\bot$ if one or both conjuncts receive $\bot$, and outputs the highest subscripted ρ_α such that one of the conjuncts receives that ρ_α otherwise. Intuitively, a conjunction where no conjunct is false and some conjunct fails to be true receives the most 'recent' pathological value had by one of its conjuncts (infinitary conjunctions involve extending this treatment in the obvious manner, remembering the observation regarding the use of *supremum* noted above):

$\wedge$	$\top$	$\bot$	ρ_1	ρ_2	ρ_3	$\cdots$
$\top$	$\top$	$\bot$	ρ_1	ρ_2	ρ_3	$\cdots$
$\bot$	$\bot$	$\bot$	$\bot$	$\bot$	$\bot$	$\cdots$
ρ_1	ρ_1	$\bot$	ρ_1	ρ_2	ρ_3	$\cdots$
ρ_2	ρ_2	$\bot$	ρ_2	ρ_2	ρ_3	$\cdots$
ρ_3	ρ_3	$\bot$	ρ_3	ρ_3	ρ_3	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Each of the relativized conditionals $\rightarrow_\alpha$, intuitively, behaves like an 'at least as true as' connective on the 'relevant' set of semantic values $\{\top, \bot\} \cup \{\rho_\beta : \beta < \alpha\}$ (where pathological values are 'less-true-than' $\bot$, and ρ_{β_1} is 'less-true-than' ρ_{β_2} if $\beta_1 > \beta_2$), and takes the 'greatest' (i.e., most recent) ρ_β taken by either the antecedent or consequent if the antecedent, consequent, or both take values outside of the relevant values $\{\top, \bot\} \cup \{\rho_\beta : \beta < \alpha\}$. Thus, the truth table for $\rightarrow_2$ is:

$\rightarrow_2$	$\top$	$\bot$	ρ_1	ρ_2	ρ_3	$\cdots$
$\top$	$\top$	$\bot$	$\bot$	ρ_2	ρ_3	$\cdots$
$\bot$	$\top$	$\top$	$\bot$	ρ_2	ρ_3	$\cdots$
ρ_1	$\top$	$\top$	$\top$	ρ_2	ρ_3	$\cdots$
ρ_2	ρ_2	ρ_2	ρ_2	ρ_2	ρ_3	$\cdots$
ρ_3	ρ_3	ρ_3	ρ_3	ρ_3	ρ_3	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

[21] Of course, the expressive completeness result proven in §6 shows that analogues of the weak Kleene connectives can also be defined within the framework.

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

and the truth table for $\rightarrow_3$ is:

$\rightarrow_3$	$\top$	$\bot$	ρ_1	ρ_2	ρ_3	$\cdots$
$\top$	$\top$	$\bot$	$\bot$	$\bot$	ρ_3	$\cdots$
$\bot$	$\top$	$\top$	$\bot$	$\bot$	ρ_3	$\cdots$
ρ_1	$\top$	$\top$	$\top$	$\bot$	ρ_3	$\cdots$
ρ_2	$\top$	$\top$	$\top$	$\top$	ρ_3	$\cdots$
ρ_3	ρ_3	ρ_3	ρ_3	ρ_3	ρ_3	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Of course, these are not the only primitive connectives and operators we could have chosen.[22] There are many others – in particular, other formulations of conjunction and the conditional—that would be just as legitimate. The only constraint in play here is that the connectives must be *monotonic* in the relevant sense. In order to make the relevant sense clear, we first need to define a class of partial orders on semantic values:

DEFINITION 3.6. *The α-ordering (α any ordinal) on:*

$$\{\top, \bot\} \cup \{\rho_\beta : 1 \leq \beta \text{ and } \beta \in ON\}$$

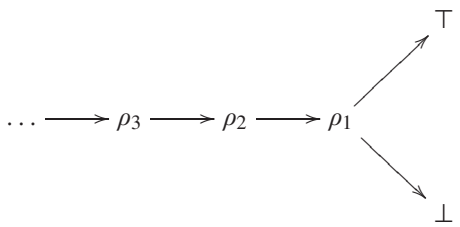
is the (unique) partial ordering $\leq_\alpha$ such that $x \leq_\alpha y$ if and only if either:

$x = \rho_{\beta_1}$, $y = \rho_{\beta_2}$, *and* $\alpha \leq \beta_2 \leq \beta_1$.

or:

$x = \rho_{\beta_1}$ *where* $\beta_1 \geq \alpha$ *and either* $y = \top$, *or* $y = \bot$, *or* $y = \rho_{\beta_2}$ *where* $\beta_2 < \alpha$.

As usual, a few examples will make the structure of the α-orderings clearer. The 1-ordering $\leq_1$ can be represented diagrammatically as:



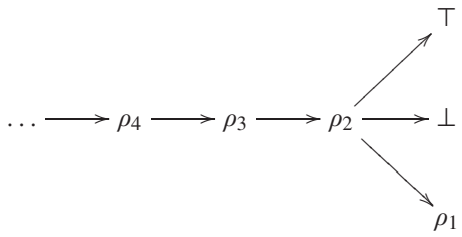
[22] In particular, there are a number of other, perhaps equally plausible, possible definitions of the relativized conditionals $\rightarrow_\alpha$. Utilizing the particular definition given above has two advantages, however (one immediately relevant, one relevant to future planned work):

1. This definition of the conditional is particularly convenient for the expressive completeness result proven in §5 below.
2. This definition of the conditional is particularly interesting in one of the deductive systems that we plan to develop in later work—one that tracks the 'at least as true as' relation sketched above.

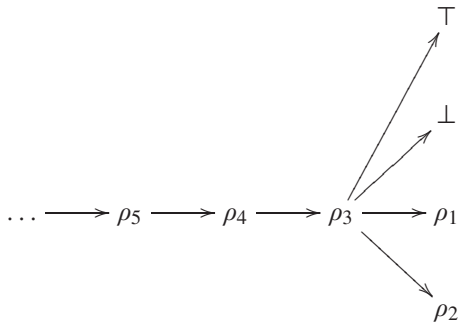
At any rate, so long as $\rightarrow_\alpha$ as defined above is a legitimate, meaningful connective, regardless of whether it really captures *the conditional* in some important sense, there should be no objection, since the expressive completeness result guarantees that the 'genuine' conditional or conditionals (whatever that turns out to be) is or are present in the formal system developed here (even if not primitive).

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

The 2-ordering $\leq_2$ can be represented diagrammatically as:



And the 3-ordering $\leq_3$ can be represented diagrammatically as:



Each connective (or, more carefully, each function on semantic values) is, or is not, monotonic relative to a particular α-ordering:

DEFINITION 3.7. *A β-ary function:*[23]

$$f : (\{\top, \bot\} \cup \{\rho_\alpha : 1 \leq \alpha \text{ and } \alpha \in ON\})^\beta \to \{\top, \bot\} \cup \{\rho_\alpha : 1 \leq \alpha \text{ and } \alpha \in ON\}$$

is α-monotonic if and only if, for any β-ary sequences of ordinals:

$$A_1 = \langle \gamma_1^1, \gamma_1^2, \gamma_1^3, \cdots \gamma_1^{\theta-1}, \gamma_1^\theta, \gamma_1^{\theta+1}, \cdots \rangle$$
$$A_2 = \langle \gamma_2^1, \gamma_2^2, \gamma_2^3, \cdots \gamma_2^{\theta-1}, \gamma_2^\theta, \gamma_2^{\theta+1}, \cdots \rangle$$

if, for any $\theta < \beta$, $\gamma_1^\theta \leq_\alpha \gamma_2^\theta$, then $f(A_1) \leq_\alpha f(A_2)$.

The following lemma is evident (and is a natural generalization of similar results for the finitary case, see (Cook 2009) and (Schlenker 2010)):

LEMMA 3.8. *For any $\beta \in ON$, all connectives in $\mathcal{L}_\beta$ are β-monotonic.*

Proof. By induction on well-formed formulas, details left to reader. □

[23] Here β can be any ordinal, and hence f can be any function from well-ordered set-sized sequences of values from our collection of semantic 'values':

$$\{\top, \bot\} \cup \{\rho_\alpha : 1 \leq \alpha \text{ and } \alpha \in ON\}$$

to single semantic values.

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

COROLLARY 3.9. *For any $\beta_1, \beta_2 \in ON$, if $\beta_1 < \beta_2$, then all connectives in $\mathcal{L}_{\beta_1}$ are β_2-monotonic.*

In order to prove the relevant generalized version of the fixed-point theorem from (Kripke 1975), we need the following 'jump' operation[24] on assignment functions $\sigma_{\mathcal{L}_\alpha}$:

DEFINITION 3.10. *Given a language $\mathcal{L}_\alpha$, a denotation function $\delta_{\mathcal{L}_\alpha}$ on $\mathcal{L}_\alpha$, and an assignment $\sigma_{\mathcal{L}_\alpha}$ on $\mathcal{L}_\alpha$, we define the* jump-successor *of $\sigma_{\mathcal{L}_\alpha}$ as follows. For any sentence name $S_\beta \in \mathcal{L}_\alpha$:*

$$\sigma^{js}_{\mathcal{L}_\alpha}(S_\beta) = \sigma^{*}_{\mathcal{L}_\alpha}(\delta_{\mathcal{L}_\alpha}(S_\beta))$$
$$\sigma^{js}_{\mathcal{L}_\alpha}(P_\beta) = \sigma^{*}_{\mathcal{L}_\alpha}(P_\beta).$$

In short, the jump-successor operation, applied to an assignment $\sigma_{\mathcal{L}_\alpha}$, generates a new assignment $\sigma^{js}_{\mathcal{L}_\alpha}$ by assigning to each sentence name S_α the semantic value had by the denotation of S_α according to the valuation $\sigma^{*}_{\mathcal{L}_\alpha}$ corresponding to the original assignment $\sigma_{\mathcal{L}_\alpha}$ (and leaving the semantic value of propositional letters unchanged). We abbreviate repeated application of the jump-successor operation as follows:

DEFINITION 3.11. *$\sigma^{js_0}_{\mathcal{L}_\alpha} = \sigma_{\mathcal{L}_\alpha}$, $\sigma^{js_{\gamma+1}}_{\mathcal{L}_\alpha} = (\sigma^{js_\gamma}_{\mathcal{L}_\alpha})^{js}$, and $\sigma^{js_\gamma}_{\mathcal{L}_\alpha}$ (γ a limit ordinal) is the unique σ such that, for any sentence name S_β in $\mathcal{L}_\alpha$:*

$$\sigma(S_\beta) = max_{\leq_\alpha}\{\sigma^{js_\rho}_{\mathcal{L}_\alpha}(S_\beta) : \rho < \gamma\},$$

where $max_{\leq_\alpha}$ is the maximum relative to the α-ordering.[25]

The following bit of notation will be helpful:

DEFINITION 3.12. *Given an expression Φ and language $\mathcal{L}_\alpha$, Φ is* initially in $\mathcal{L}_\alpha$ *($\Phi \in_I \mathcal{L}_\alpha$) if and only if $\Phi \in \mathcal{L}_\alpha$ and there is no $\beta < \alpha$ such that $\Phi \in \mathcal{L}_\beta$.*

We can now prove the existence of fixed-points for each language $\mathcal{L}_\alpha$ along familiar lines.

LEMMA 3.13. *Given any denotation function $\delta_{\mathcal{L}_\alpha}$ on $\mathcal{L}_\alpha$, let $\sigma_{\mathcal{L}_\alpha}$ be any assignment such that:*

$$\sigma_{\mathcal{L}_\alpha}(S_\beta) = \rho_\gamma,$$

where γ is the unique ordinal such that $\delta_{\mathcal{L}_\alpha}(S_\beta) \in_I \mathcal{L}_\gamma$. Then there is a fixed point *in the sequence of jump-successors of $\sigma_{\mathcal{L}_\alpha}$. That is, there is an ordinal θ such that:*

$$\sigma^{js_\theta}_{\mathcal{L}_\alpha} = \sigma^{js_{\theta+1}}_{\mathcal{L}_\alpha}$$

[24] Since we are, from a mathematical perspective, working with infinitely-many (in fact, proper-class-many) semantic values at once, it is more convenient to formulate the 'jump' operation on valuations rather than on extensions, anti-extensions, etc., of the truth predicate. The end result, however, is similar.

[25] Note that the α-monotonicity of all connectives in $\mathcal{L}_\alpha$ guarantees that there will be a unique such maximum in the cases of interest, even if this is not always so for arbitrary functions.

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

Proof. Straightforward generalization of (Kripke 1975), given the α-monotonicity of all connectives in $\mathcal{L}_\alpha$ and the fact that $\mathcal{L}_\alpha$ is set-sized. See (Cook 2008), (Schlenker 2010) for further details.[26] □

THEOREM 3.14. *Given any denotation function $\delta_{\mathcal{L}_\alpha}$ on $\mathcal{L}_\alpha$, let $\sigma_{\mathcal{L}_\alpha}$ be any assignment such that:*

$$\sigma_{\mathcal{L}\alpha}(S_\beta) = \rho_\gamma,$$

where γ is the unique ordinal such that $\delta_{\mathcal{L}_\alpha}(S_\beta) \in_I \mathcal{L}_\gamma$. Then (with θ as in previous result):

$$\mathcal{M} = \langle \delta_{\mathcal{L}_\alpha}, \sigma_{\mathcal{L}_\alpha}^{js_\theta} \rangle$$

is a model (the minimal fixed point model*) of $\mathcal{L}_\alpha$.*

Proof. Straightfoward consequence of Lemma 3.13 and the definition of acceptable assignment. □

Thus, we can, via a straightforward (albeit transfinite) generalization of standard and well-known constructions, provide models for each language $\mathcal{L}_\alpha$ in our indefinitely extensible hierarchy of languages. Of course, there is much more that can (and should) be proven and said about these models and their properties (for example, there is much to be said about the relation between the minimal fixed points constructed above and various other fixed points, such as maximal intrinsic fixed points). We plan to return to this topic in further work (or see (Cook 2008) and (Schlenker 2010) for relevant results proven for similar, but somewhat expressively impoverished, systems).

§4. Other Fixed-point Approaches. In the sections that follow we shall concentrate on the formal semantics given in the previous section. The reason for this narrow focus is straightforward: The generalization of (Kripke 1975) developed in (Cook 2008), (Cook 2009), (Schlenker 2010), and the previous section is, we believe, a simple and natural extension of Kripke's original three-valued construction. All that is involved, viewed from a purely mathematical perspective, is replacing the single pathological value ρ with an indefinitely extensible sequence of pathological values:

$$\rho_1, \rho_2, \rho_3 \ldots \rho_\omega, \rho_{\omega+1}, \ldots$$

identifying appropriate orderings on these values, and iterating the fixed point construction relative to these orderings.

It is worth noting, however, that there are other ways that Kripke's construction has been expanded or modified. One that is of particular interest here stems from an idea originally due to JC Beall in (Beall 2006).[27] Beall suggests that we can coherently add

[26] The complicated condition assigning different ρ_γ to distinct sentence letters, rather than simply assigning the most 'recent' such value ρ_α, is required in order to guarantee that, for any languages $\mathcal{L}_{\beta_1}$ and $\mathcal{L}_{\beta_2}$ where $\beta_1 < \beta_2$, denotation functions $\delta_{\mathcal{L}_{\beta_1}}$ and $\delta_{\mathcal{L}_{\beta_2}}$ where, for all $S_\gamma \in \mathcal{L}_{\beta_1}$, $\delta_{\mathcal{L}_{\beta_1}}(S_\gamma) = \delta_{\mathcal{L}_{\beta_2}}(S_\gamma)$, and sentence $\Phi \in \mathcal{L}_{\beta_1}$, Φ will receive the same semantic value on the minimal fixed points for $\mathcal{L}_{\beta_1}$ and $\mathcal{L}_{\beta_2}$.

[27] Beall introduces the terminology "paranormal" in his description of the proposed alteration to Kripke's original construction. We shall continue to use "pathological", however, both to maintain terminological consistency and because (Rosenblatt & Szmuc 2014) also use it.

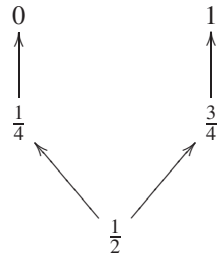
https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

a pathologicality operator to our language—that is, an operator π such that:

$$\begin{aligned}\sigma(\pi(\Phi)) &= 1 \text{ iff } \sigma(\Phi) = \rho.\\ &= 0 \text{ iff } \sigma(\Phi) \neq \rho,\end{aligned}$$

by allowing truth, falsity, and pathologicality to overlap (while not necessarily allowing truth and falsity to overlap), resulting in a semantics that categorizes sentences in terms of five categories: true simpliciter, false simpliciter, true and pathological, false and pathological, and pathological simpliciter.

Lucas Rosenblatt and Damián Szmuc develop a Kripkean fixed-point semantics corresponding to this idea in (Rosenblatt & Szmuc 2014). The technical details are complicated and subtle, but the basic idea is as follows: First, we introduce five semantic values $\{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$ where 1 represents truth simpliciter, 0 represents falsity simpliciter, $\frac{3}{4}$ represents pathological truth, $\frac{1}{4}$ represents pathological falsity, and $\frac{1}{2}$ represents pathologicality simpliciter. They then introduce the following order on these values:



Finally, they introduce a Kripkean jump operator for the pathologicality operator π that is monotonic with respect to this ordering, hence guaranteeing that the construction will have a fixed point.[28] Interestingly, the value that $\pi(\Phi)$ receives at a particular stage of the construction does not depend solely on the value assigned to Φ at that stage, but also depends on values assigned to Φ on valuations that *extend* the valuation of that particular stage. In short, the pathologicality operator π is not truth-functional on this approach.[29]

---

[28] It is for this reason that we concentrate on (Rosenblatt & Szmuc 2014), rather than on (Beall 2006), since the operator introduced in (Beall 2006) is not monotonic with respect to this ordering (or any other appropriate ordering). See (Rosenblatt & Szmuc 2014): 604–605.

[29] The non-truth-fuctionality of the pathological operator π in (Rosenblatt & Szmuc 2014) also explains how *Theorem 1.4* is compatible with the fact that their view stably assigns a semantic value to a sentence λ_S where λ_S is equivalent to:

$$\pi(\ulcorner \lambda_S \urcorner)$$

Theorem 1.4 implies that, if the collection of semantic values is finite (hence definite), then there can be no predicate Φ_S such that:

$$\begin{aligned}\sigma(\Phi_S(\ulcorner \Psi \urcorner)) &= 1 \text{ iff } \sigma(\Psi) \neq \top.\\ &= 0 \text{ otherwise.}\end{aligned}$$

One might think that the view developed in (Rosenblatt & Szmuc 2014) contains exactly such a predicate—namely:

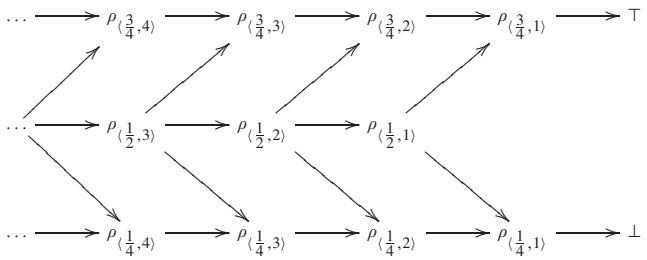
$$\pi(T(\ulcorner x \urcorner)) \vee \neg T(\ulcorner x \urcorner).$$

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

We need not belabor any more of the (interesting and sophisticated) details of Rosenblatt and Szmuc's construction, since we have no intention of proving any theorems about this alternative approach here. Instead, we wish merely to use (Rosenblatt & Szmuc 2014) to clarify an important point: the fact that the Embracing Revenge account is a general approach, rather than a particular formalization. In other words, the technical details of the Embracing Revenge account are flexible and negotiable. There are various ways that we might modify or complicate the semantics described in the previous section. For example, we could replace (our generalization of) the strong Kleene account of the connectives with (a corresponding generalization of) the weak Kleene connectives, or we could take the the maximal intrinsic fixed point rather than the minimal fixed point at each stage in the construction. Rosenblatt and Szmuc's construction shows that there is another way we could modify the view: we could apply the general idea underlying the Embracing Revenge view, not to Kripke's original construction (replacing one pathological value with many), but rather to one or another of the recent modifications or extensions of Kripke's account. Doing so in the specific case of (Rosenblatt & Szmuc 2014) would require us to:

1. Replace the finite set of pathological values with a transfinite sequence of pathological values, and define an appropriate ordering on them.
2. Provide semantic clauses for each of the operators and predicates in the language.[30]
3. Prove that these operators and predicates are monotonic with respect to the ordering in question.
4. Prove an expressive completeness result for the new system analogous to *Theorems 6.5* and *6.6* below.

Space considerations prevent us from exploring the much more complicated construction that would result from such an approach.[31] But we can sketch how such an account might begin. Just as we replaced Kripke's single pathological value with an infinite, indefinitely extensible sequence of pathological values in the account above, we could likewise replace the three pathological values found in (Rosenblatt & Szmuc 2014) with three indefinitely extensible sequences of pathological values. As a result, we would obtain the following alternative 1_{RS}-ordering:



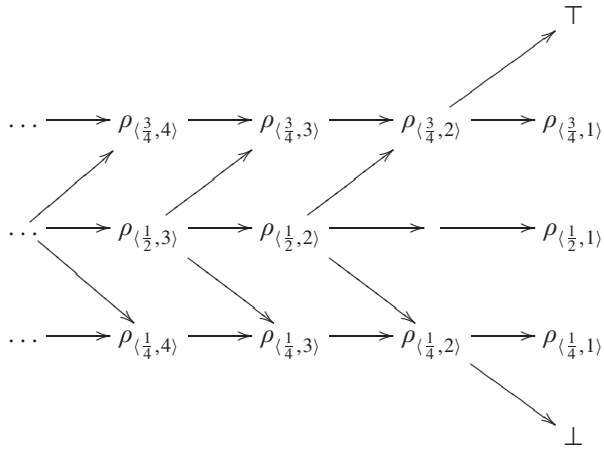
But the non-truth-functionality of π prevents this complex predicate from behaving in the way prohibited by **Theorem 1.4**. Any sentence λ equivalent to the predication of this predicate to λ turns out pathological and true—that is, receives value $\frac{3}{4}$—according to (Rosenblatt & Szmuc 2014)).

30 Which, on the approach being imagined here, would presumably include all of the operators and predicates of the language(s) described in the previous section, plus Rosenblatt and Szmuc's novel nontruth functional operator π.

31 We do plan on exploring this more complicated construction in detail in future work, however.

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

The 2_{RS}-ordering would look like:



And so on. We would then apply generalizations of the treatments of the operators—including the non-truth-functional pathologicality operator π—to these orderings, in order to obtain an indefinitely extensible series of fixed points.

The system that results would be even more expressively rich than the formal system sketched in §3 above, and this is no doubt a good thing. But we will not explore such extensions here. Instead, we will (for the purposes of this paper, at least) rest content with noting that this brief discussion of this more complicated approach illuminates two important aspects of the present approach:

- The preliminary sketch of a Rosenblatt and Szmuc-style variant of the Embracing Revenge approach demonstrates that Embracing Revenge is a broad technique for dealing with the Revenge problem, and can be adapted to very different accounts of the exact semantics of the connectives and quantifiers (so long as they are broadly Kripkean in nature).[32]
- The additional expressive resources provided by the inclusion of a pathologicality operator of the sort explored by Rosenblatt and Szmuc are not needed in order to prove the expressive completeness result given in §6 below.[33]

Now that we have seen, in at least broad outline, how to generalize the Embracing Revenge view to alternative, modified finitary versions of the construction found in (Kripke 1975), it is time to return to the simple semantics developed in detail in the previous section and explore some other, defined notions that will be useful in the expressive completeness proof to follow.

[32] For example, we could also apply the general Embracing Revenge framework to the semantics developed in (Field 2008). The resulting account, while no doubt *extremely* complex, would increase the already impressive expressive power of Field's account. In particular, it would allow for the formulation of additional well behaved determinacy operators inexpressible in Field's official account (such as D^{∞} operators, which hold of a formula if and only if D^{α} holds for all ordinals α).

[33] The reason is simple: The expressive completeness result, briefly put, shows that any *extensional* notion can be expressed at some level of the hierarchy of languages. As is discussed at length in (Rosenblatt & Szmuc 2014), the pathologicality operator is *intensional* as a result of its non-truth-functional character.

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

§5. Other Connectives and Predicates: Examples. The first, and in some ways most important, defined connective (given that the project undertaken here is motivated by the Liar paradox and revenge-style variants of it) is the transparent falsity predicate, which is defined as follows:[34]

$$F(x) =_{df} \neg T(x)$$

The following table summarizes the logical behavior of our falsity predicate.

S_α	$F(S_\alpha)$
$\top$	$\bot$
$\bot$	$\top$
ρ_1	ρ_1
ρ_2	ρ_2
$\vdots$	$\vdots$
ρ_β	ρ_β
$\vdots$	$\vdots$

Note that, like the truth predicate, the falsity predicate maps each pathological value ρ_α to itself—hence, the slight abuse of jargon: the transparent falsity predicate is transparent on the nonclassical semantic values.

The next defined connective is disjunction, which can be obtained using conjunction and negation via the standard DeMorgan-style definition:

$$\Phi \vee \Psi =_{df} \neg(\neg\Phi \wedge \neg\Psi)$$
$$\vee(\Sigma) =_{df} \neg(\wedge(\{\neg\Phi : \Phi \in \Sigma\})).$$

The binary version of disjunction outputs $\bot$ if both disjuncts receive $\bot$, outputs $\top$ if one or both disjuncts receive $\top$, and outputs the highest subscripted ρ_α such that one of the disjuncts receives that ρ_α otherwise. Intuitively, like conjunction, a disjunction where no disjunct is true and some disjunct fails to be false receives the most 'recent' pathological value had by one of its disjuncts (infinitary disjunctions involve extending this treatment in the obvious manner):

$\vee$	$\top$	$\bot$	ρ_1	ρ_2	ρ_3	$\cdots$
$\top$	$\top$	$\top$	$\top$	$\top$	$\top$	$\cdots$
$\bot$	$\top$	$\bot$	ρ_1	ρ_2	ρ_3	$\cdots$
ρ_1	$\top$	ρ_1	ρ_1	ρ_2	ρ_3	$\cdots$
ρ_2	$\top$	ρ_2	ρ_2	ρ_2	ρ_3	$\cdots$
ρ_3	$\top$	ρ_3	ρ_3	ρ_3	ρ_3	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

[34] Note that we cannot define the falsity predicate as:

$$F(x) =_{df} T(\neg x)$$

since the negation operator does not apply directly to sentence names (which are the only expressions that can occur as arguments of the truth predicate).

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

Given the presence of both conjunction and the transfinite sequence of relativized conditionals, it is natural to introduce a sequence of defined relativized biconditionals:

$$(\Phi \leftrightarrow_\beta \Psi) =_{df} ((\Phi \rightarrow_\beta \Psi) \wedge (\Psi \rightarrow_\beta \Phi)).$$

An important reminder: For each ordinal α, the relativized conditional $\rightarrow_\alpha$ only occurs in languages $\mathcal{L}_\beta$ such that $\beta \geq \alpha$, and as a result, each relativized biconditional $\leftrightarrow_\alpha$, since defined in terms of $\rightarrow_\alpha$, only occurs in languages $\mathcal{L}_\beta$ such that $\beta \geq \alpha$. This holds generally in what follows: for any relativized connective $\otimes_\alpha$ introduced below, $\otimes_\alpha$ occurs in all and only those languages $\mathcal{L}_\beta$ where $\beta \geq \alpha$.

A relativized biconditional $\leftrightarrow_\alpha$, intuitively, behaves like a 'has the same semantic value as' connective on the 'relevant' set of semantic values $\{\top, \bot\} \cup \{\rho_\beta : \beta < \alpha\}$, and takes the 'greatest' (i.e., most recent) ρ_β taken by either the antecedent or consequent if the antecedent, consequent, or both take values outside of the relevant values $\{\top, \bot\} \cup \{\rho_\beta : \beta < \alpha\}$. Thus, the truth table for $\leftrightarrow_2$ is:

$\leftrightarrow_2$	$\top$	$\bot$	ρ_1	ρ_2	ρ_3	$\cdots$
$\top$	$\top$	$\bot$	$\bot$	ρ_2	ρ_3	$\cdots$
$\bot$	$\bot$	$\top$	$\bot$	ρ_2	ρ_3	$\cdots$
ρ_1	$\bot$	$\bot$	$\top$	ρ_2	ρ_3	$\cdots$
ρ_2	ρ_2	ρ_2	ρ_2	ρ_2	ρ_3	$\cdots$
ρ_3	ρ_3	ρ_3	ρ_3	ρ_3	ρ_3	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

and the truth table for $\leftrightarrow_3$ is:

$\leftrightarrow_3$	$\top$	$\bot$	ρ_1	ρ_2	ρ_3	$\cdots$
$\top$	$\top$	$\bot$	$\bot$	$\bot$	ρ_3	$\cdots$
$\bot$	$\bot$	$\top$	$\bot$	$\bot$	ρ_3	$\cdots$
ρ_1	$\bot$	$\bot$	$\top$	$\bot$	ρ_3	$\cdots$
ρ_2	$\bot$	$\bot$	$\bot$	$\top$	ρ_3	$\cdots$
ρ_3	ρ_3	ρ_3	ρ_3	ρ_3	ρ_3	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

As noted above, we cannot, within the Embracing Revenge framework, have a genuine weak negation $\neg_W$ (i.e., a genuine 'fails to be true' connective that outputs $\bot$ when the input is $\top$ and outputs $\top$ otherwise) or a genuine strong negation $\neg_S$ (i.e., a genuine 'is the false' connective that outputs $\top$ when the input is $\bot$ and outputs $\bot$ otherwise).[35] In short,

[35] This is as it should be, since a genuine weak negation is equivalent to an operator expressing the notion:

> Has one of *the* truth values (i.e., *all* of them) other than the true.

and a genuine strong negation is equivalent to an operator expressing the notion:

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

we cannot have either of:

Φ	$\neg_W \Phi$	$\neg_S \Phi$
$\top$	$\bot$	$\bot$
$\bot$	$\top$	$\top$
ρ_1	$\top$	$\bot$
ρ_2	$\top$	$\bot$
$\vdots$	$\vdots$	$\vdots$
ρ_β	$\top$	$\bot$
$\vdots$	$\vdots$	$\vdots$

We can, however, introduce a sequence of relativized weak negations defined as follows:

$$\begin{aligned}\neg^W_\beta \Phi &=_{df} \Phi \rightarrow_\beta C_0\\ &=_{df} \Phi \rightarrow_\beta \bot.\end{aligned}$$

Each relativized weak negation $\neg^W_\beta$ outputs $\bot$ if the input is $\top$, outputs $\top$ if the input is $\bot$ or ρ_γ where $0 < \gamma < \beta$, and acts as the unary identity function otherwise. Thus:

Φ	$\neg^W_1 \Phi$	$\neg^W_2 \Phi$	$\neg^W_3 \Phi$	$\neg^W_4 \Phi$	$\cdots$
$\top$	$\bot$	$\bot$	$\bot$	$\bot$	$\cdots$
$\bot$	$\top$	$\top$	$\top$	$\top$	$\cdots$
ρ_1	ρ_1	$\top$	$\top$	$\top$	$\cdots$
ρ_2	ρ_2	ρ_2	$\top$	$\top$	$\cdots$
ρ_3	ρ_3	ρ_3	ρ_3	$\top$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Along similar lines, we can introduce a tranfinite sequence of relativized strong negations as follows:[36]

$$\begin{aligned}\neg^S_\beta \Phi &= \neg\neg^W_\beta \neg\Phi\\ &= \neg(\neg\Phi \rightarrow_\beta C_0)\\ &= \neg(\neg\Phi \rightarrow_\beta \bot).\end{aligned}$$

Fails to have *any* truth values (i.e., any of *all* of them) other than the false.

Both of these have already been ruled out: in a technical sense, by Theorem 1.4, and in a philosophical sense by our rejection of absolutely general quantification.

[36] Equivalently, we could have defined $\neg^S_\beta \Phi$ as $\neg(\neg\Phi \rightarrow_\beta \bot)$, and then defined $\neg^W_\beta \Phi$ as $\neg\neg^S_\beta \neg\Phi$. As a result, we obtain the following elegant equivalences:

$$\begin{aligned}\neg\neg^S_\beta \Phi &\equiv \neg^W_\beta \neg\Phi\\ \neg^S_\beta \neg\Phi &\equiv \neg\neg^W_\beta \Phi.\end{aligned}$$

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

Each relativized strong negation $\neg^S_\beta$ outputs $\top$ if the input is $\bot$, outputs $\bot$ if the input is $\top$ or ρ_γ where $0 < \gamma < \beta$, and acts as the unary identity function otherwise. Thus:

Φ	$\neg^S_1\Phi$	$\neg^S_2\Phi$	$\neg^S_3\Phi$	$\neg^S_4\Phi$	$\cdots$
$\top$	$\bot$	$\bot$	$\bot$	$\bot$	$\cdots$
$\bot$	$\top$	$\top$	$\top$	$\top$	$\cdots$
ρ_1	ρ_1	$\bot$	$\bot$	$\bot$	$\cdots$
ρ_2	ρ_2	ρ_2	$\bot$	$\bot$	$\cdots$
ρ_3	ρ_3	ρ_3	ρ_3	$\bot$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

We next introduce a series of relativized truth predicates. In a multi-valued context there are (at least) two kinds of truth predicate. The first is a *transparent* truth predicate—a predicate $T(x)$ such that the result of applying this predicate to the name of a sentence Φ should receive the same semantic value as Φ itself. Such a predicate is, of course, a primitive notion in the present system (and we have already seen the defined 'transparent' falsity predicate $F(x)$). Intuitively, there is a second notion—what we shall call here the *strict truth predicate*—where application of this sort of truth predicate to a sentence Φ—that is:

$$\Phi \text{ is (strictly) true}$$

should be true if Φ is true, and false otherwise (i.e., if Φ receives any value *other* than $\top$, including, in the present context, any of the ρ_αs). Such a predicate would not be monotonic relative to any of the α-orderings $\leq_\alpha$, however.[37] We can define a sequence of relativized truth predicates $T_\beta(x)$ that 'approximate' the strict truth predicate (in roughly the same sense that the relativized conditionals $\rightarrow_\beta$ 'approximate' the intuitive, but impossible, reading of the material conditional):[38]

$$\begin{aligned} T_\beta(x) &=_{df} \neg^W_\beta \neg^W_\beta T(x). \\ &=_{df} ((T(x) \rightarrow_\beta C_0) \rightarrow_\beta C_0) \\ &=_{df} ((T(x) \rightarrow_\beta \bot) \rightarrow_\beta \bot) \end{aligned}$$

[37] This is a bit loose, since we have only officially defined monotonicity for connectives, not for predicates. The point can be put a bit more precisely as follows: We can easily define a corresponding notion of monotonicity on predicates analogous to the notion as applied to connectives above. Given that the only primitive predicate in the formal system is the truth predicate "$T(x)$" (which is a kind of identity function, mapping each sentence name to a sentence with the same semantic status), we can easily prove that all predicates definable within any language $\mathcal{L}_\alpha$ are α-monotonic in this generalized sense via a straightforward induction. The strict notion of truth would not be α-monotonic for any ordinal α, however.

[38] The definition of the relativized truth predicate suggests a natural doubly-subscripted generalization:

$$T_{\langle\beta_1,\beta_2\rangle}(x) =_{df} ((T(x) \rightarrow_{\beta_1} \bot). \rightarrow_{\beta_2} \bot)$$

We plan to explore the behavior of this interesting notion in future work.

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

The following table summarizes the logical behavior of these relativized strict truth predicates:

S_γ	$T_1(S_\gamma)$	$T_2(S_\gamma)$	$T_3(S_\gamma)$	$T_4(S_\gamma)$	$\cdots$
$\top$	$\top$	$\top$	$\top$	$\top$	$\cdots$
$\bot$	$\bot$	$\bot$	$\bot$	$\bot$	$\cdots$
ρ_1	ρ_1	$\bot$	$\bot$	$\bot$	$\cdots$
ρ_2	ρ_2	ρ_2	$\bot$	$\bot$	$\cdots$
ρ_3	ρ_3	ρ_3	ρ_3	$\bot$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

We can also define relativized falsity predicates along the following lines:

$$F_\beta(x) =_{df} \neg^W_\beta \neg^W_\beta F(x).$$
$$=_{df} ((F(x) \rightarrow_\beta C_0) \rightarrow_\beta C_0).$$

The following table displays the behavior of these predicates:

S_γ	$F_1(S_\gamma)$	$F_2(S_\gamma)$	$F_3(S_\gamma)$	$F_4(S_\gamma)$	$\cdots$
$\top$	$\bot$	$\bot$	$\bot$	$\bot$	$\cdots$
$\bot$	$\top$	$\top$	$\top$	$\top$	$\cdots$
ρ_1	ρ_1	$\bot$	$\bot$	$\bot$	$\cdots$
ρ_2	ρ_2	ρ_2	$\bot$	$\bot$	$\cdots$
ρ_3	ρ_3	ρ_3	ρ_3	$\bot$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

The relativized strict truth and falsity predicates are of particular interest since they strikingly illustrate an additional motivation—one distinct from, although intimately related to, the indefinite extensibility concerns argued for in (Cook 2008) and (Cook 2009), and rehearsed above—for allowing *only* the relativized versions of the strict notions of truth and falsity, while disallowing the absolute, nonrelativized (and non-α-monotonic, for all $\alpha \in ON$) strict truth and falsity predicate.

(Schlenker 2010) argues that it is in fact the relativized versions of strict truth and strict falsity that are typically used when asserting that sentences—especially sentences such as the Liar and its strengthened cousins that take pathological semantic values—are or are not true. With respect to the relativized falsity predicate he writes that:

> The expression *F is something other than true* or *F has a value different from true* offer particularly clear expressions... In both cases, the negative expression includes a quantifier over truth values. A proper semantic analysis must take this into account, and derive the meaning of these expressions from independent properties of natural language quantification. At the same time, however, we must recognize that the paradoxes force speakers to reconsider the assumptions they may normally make about the domain of possible truth values. One way to interpret our theory is to assume that a speaker who reflects on the paradoxical statements will *gradually expand the domain of truth values that he considers*... this

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

> fact is accommodated by positing that the implicit domain restriction on the quantifier *something other than true* (which for our purposes is equivalent to *has a value different from true*) will change as the speaker reflects on paradoxes. ((Schlenker 2010): 377–378)[39]

In short, when first confronted with the $\mathcal{L}_1$ Liar sentence λ_1 (notation as in §1), we are (rightly) drawn to conclude that λ_1 is *not true*. This claim, however, is an $\mathcal{L}_2$ assertion of the form $\neg_2^W T(\lambda_1)$—that is, the claim that λ_1 has a value other than true with respect to the semantic values currently 'in play', so to speak: $\top$, $\bot$, and ρ_1.[40] Note, however, that:

$$\neg_2^W T(\Phi) \equiv \neg T_2(\Phi).$$

Likewise, when confronted with the strengthened Liar sentence λ_2, we are also (rightly) drawn to conclude that λ_2 is *not true*. This claim, however, is an $\mathcal{L}_3$ assertion of the form $\neg_3^W T(\lambda_2)$—that is, the claim that λ_2 has a value other than true with respect to the truth values *now* currently 'in play': $\top$, $\bot$, ρ_1, and ρ_2. Again, note that:[41]

$$\neg_3^W T(\Phi) \equiv \neg T_3(\Phi).$$

Thus, in moving from consideration of the Liar λ_1 to consideration of the strengthened Liar λ_2, we are forced to expand the range of semantic values that are relevant to semantic assessments. More generally, as we successively encounter 'higher-level' instances of Revenge (or higher-level expressions of any sort), the domain of semantic values that are relevant expands, and our judgements that these sentences are *not true* involve more expansive (in terms of the relevant semantic values 'in play') relativized strict truth predicates (or, equivalently, more expansive relativized weak negations).

The final defined notion we shall introduce in this section is a sequence of relativized γ-pathological operators. As we noted in the introduction, we cannot allow genuine γ-pathological operators $C^\gamma(\cdots)$ such that $C^\gamma(\cdots)$ outputs true when the argument takes ρ_γ as semantic value, and outputs false otherwise (in technical terms, such connectives would not be α-monotonic for any α). In short, we cannot introduce connectives that behave

---

[39] The ellipsis in this quotation is due to the fact that this passage of (Schlenker 2010) actually addresses the sequence of relativized negations taken to be primitive in that work—negations equivalent to the weak negations $\neg_\beta^W$ introduced above. Schlenker concludes the passage by noting that similar considerations apply to the alternative formalization presented later in the paper, which involves a sequence of relativized truth predicates equivalent to the $T_\beta(x)$s just defined:

> ... there might be other expressions of weak negation in English, and we will consider in section 2 another way in which a related meaning can be obtained. ((Schlenker 2010): 378)

The equivalence of addressing this issue in terms of relativized negations, and in terms of relativized truth predicates, is made clear in the treatment above.

[40] Recall that $\mathcal{L}_2$ requires the second pathological value ρ_2 in its semantics, but can only fully 'describe' the semantics of $\mathcal{L}_1$, and hence can only fully 'describe' $\top$, $\bot$, and ρ_1.

[41] Of course, once we have judged λ_2 to be *not true* in the $\neg T_2(x)$ sense, we are also free to go back and apply this new *not true* predicate, with its wider domain of relevant semantic values, to the original Liar sentence λ_1.

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

as follows:

Φ	$C^0(\Phi)$	$C^1(\Phi)$	$C^2(\Phi)$	$C^3(\Phi)$	$\cdots$
$\top$	$\bot$	$\bot$	$\bot$	$\bot$	$\cdots$
$\bot$	$\top$	$\bot$	$\bot$	$\bot$	$\cdots$
ρ_1	$\bot$	$\top$	$\bot$	$\bot$	$\cdots$
ρ_2	$\bot$	$\bot$	$\top$	$\bot$	$\cdots$
ρ_3	$\bot$	$\bot$	$\bot$	$\top$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$
ρ_β	$\bot$	$\bot$	$\bot$	$\bot$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

We can, however, introduce relativized γ-pathological operators. Each such operator has both a superscript and a subscript—the superscript marks the pathological value being 'captured', and (following the notational conventions already adopted) the subscript marks the 'first' language in which that operator is definable:[42]

$$C^\gamma_\beta(\Phi) =_{df} \Phi \leftrightarrow_\beta C_\gamma .$$

Intuitively, $C^\gamma_\beta(\cdots)$ is true of Φ if and only if Φ has semantic value ρ_γ, and false otherwise, insofar as $\mathcal{L}_\beta$ is expressive enough to make such distinctions (i.e., if $\gamma < \beta$) and receives the 'maximum' of ρ_β and the value of Φ otherwise. The following table summarizes the behavior of the sequence of relativized pathologicality operators that can be constructed using the $\mathcal{L}_3$ relativized biconditional $\leftrightarrow_3$:[43]

Φ	$C^0_3(\Phi)$	$C^1_3(\Phi)$	$C^2_3(\Phi)$	$C^3_3(\Phi)$	$C^4_3(\Phi)$	$\cdots$
$\top$	$\bot$	$\bot$	$\bot$	ρ_3	ρ_4	$\cdots$
$\bot$	$\top$	$\bot$	$\bot$	ρ_3	ρ_4	$\cdots$
ρ_1	$\bot$	$\top$	$\bot$	ρ_3	ρ_4	$\cdots$
ρ_2	$\bot$	$\bot$	$\top$	ρ_3	ρ_4	$\cdots$
ρ_3	ρ_3	ρ_3	ρ_3	ρ_3	ρ_4	$\cdots$
ρ_4	ρ_4	ρ_4	ρ_4	ρ_4	ρ_4	$\cdots$
ρ_5	ρ_5	ρ_5	ρ_5	ρ_5	ρ_5	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

[42] Of course, we could also define a sequence of analogous pathologicality predicates by inserting the truth predicate into the definition in the obvious manner:

$$P^\gamma_\beta(x) =_{df} T(x) \leftrightarrow_\beta C_\gamma .$$

[43] Note that $C^0_\gamma(\cdots)$ is logically equivalent to $\neg^S_\gamma(\cdots)$.

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

Likewise, the following table summarizes the behavior of the (loosely speaking, 'improved') sequence of pathologicality operators occurring in $\mathcal{L}_4$ (of course, the older $C_3^\gamma(\cdots)$ operators continue to be present in $\mathcal{L}_4$):

Φ	$C_4^0(\Phi)$	$C_4^1(\Phi)$	$C_4^2(\Phi)$	$C_4^3(\Phi)$	$C_4^4(\Phi)$	$\cdots$
$\top$	$\bot$	$\bot$	$\bot$	$\bot$	ρ_4	$\cdots$
$\bot$	$\top$	$\bot$	$\bot$	$\bot$	ρ_4	$\cdots$
ρ_1	$\bot$	$\top$	$\bot$	$\bot$	ρ_4	$\cdots$
ρ_2	$\bot$	$\bot$	$\top$	$\bot$	ρ_4	$\cdots$
ρ_3	$\bot$	$\bot$	$\bot$	$\top$	ρ_4	$\cdots$
ρ_4	ρ_4	ρ_4	ρ_4	ρ_4	ρ_4	$\cdots$
ρ_5	ρ_5	ρ_5	ρ_5	ρ_5	ρ_5	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Although there are many other interesting notions that can be defined within the hierarchy of languages $\mathcal{L}_\alpha$, these notions suffice for our purposes: first, to demonstrate the expressive richness of the language (and the important fact that the languages continue to get more expressive as we ascend in the hierarchy), and second, to set up those particular defined notions that will be useful in the proof of expressive completeness. Thus, we now turn to that proof.

§6. Expressive Completeness. Loosely speaking, an expressive completeness theorem for a language $\mathcal{L}$ typically takes something like the following form:

> Given any arbitrary function $\mathcal{F}$ from sequences of semantic values (of whatever sort) to semantic values (of whatever sort), there is an expression of $\mathcal{L}$ that captures $\mathcal{F}$.

Of course, we shall have to tighten this up a bit before we can do any serious mathematics with the notion. In particular, there are two critical notions in this rough version that need to be explicated carefully: what we mean by "expression", and what we mean by "capture".

We shall start with the latter issue, since it is in a sense where the present treatment differs most from traditional proofs of expressive completeness (because, e.g., standard propositional logics only have one kind of expression). Although the languages $\mathcal{L}_\alpha$ do not contain quantifiers, they do contain both sentence names and predicates that hold of those sentence names in addition to the typical trappings of propositional logic (i.e., propositional letters and connectives). Thus, we need to ask: When attempting to capture a particular truth function (or when attempting to prove that all truth functions are able to be so captured), should it be a defined logical connective (i.e., a sentence with propositional letters acting as ersatz propositional variables), or a predicate (i.e., a sentence with sentence names acting as ersatz objectual variables), that does the relevant capturing?

It turns out that we do not have to decide—we can carry out the proof using either notion. Thus, we shall prove the version where truth functions are represented by defined connectives first, and then show how to obtain the alternative version of the result as an easy corollary.

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

Now that we know what sorts of expression are to do the 'expressing', in the relevant sense, we can provide a straightforward account of what it means for such an expression to 'capture' a particular function that maps (possibly infinite) sequences of truth values to individual truth values. First, the 'connective' reading:

DEFINITION 6.1. *Given any set of semantic values* Δ *(finite or infinite), any total* α*-ary function (*α any *ordinal)* $\mathcal{F}$ *mapping* α*-sequences of members of* Δ *to members of* Δ*, and any* $\mathcal{L}_\beta$ *sentence* Φ *all of whose propositional letters are contained in:*

$$\{P_\theta : \theta < \alpha\}.$$

$\mathcal{F}$ *is* connectively expressed by Φ *if and only if, for any* α*-ary sequence of semantic values:*

$$A = \langle \gamma_1, \gamma_2, \gamma_3, \cdots \gamma_{\theta-1}, \gamma_\theta, \gamma_{\theta+1}, \cdots \rangle,$$

where $\gamma_\theta \in \Delta$ *for all* $\gamma < \alpha$*, and any assignment* $\sigma_{\mathcal{L}_\beta}$ *such that:*

$$\sigma_{\mathcal{L}_\beta}(P_\theta) = \gamma_\theta$$

we have:

$$\mathcal{F}(A) = \sigma^*_{\mathcal{L}_\beta}(\Phi).$$

$\mathcal{F}$ *is* connectively expressible in $\mathcal{L}_\beta$ *if and only if there is a* $\Phi \in \mathcal{L}_\beta$ *such that* Φ *connectively expresses* $\mathcal{F}$.

A similar definition for the 'predicative' understanding of expressive completeness is straightforward:[44]

DEFINITION 6.2. *Given any set of semantic values* Δ *(finite or infinite), any total* α*-ary function (*α any *ordinal) and any* $\mathcal{F}$ *mapping* α*-sequences of members of* Δ *to members of* Δ*, and any* $\mathcal{L}_\beta$ *sentence* Φ *all of whose sentence names are contained in:*

$$\{S_\theta : \theta < \alpha\}$$

$\mathcal{F}$ *is* predicatively expressed by Φ *if and only if, for any* α*-ary sequence of semantic values:*

$$A = \langle \gamma_1, \gamma_2, \gamma_3, \cdots \gamma_{\theta-1}, \gamma_\theta, \gamma_{\theta+1}, \cdots \rangle$$

where $\gamma_\theta \in \Delta$ *for all* $\gamma < \alpha$*, and any assignment* $\sigma_{\mathcal{L}_\beta}$ *such that:*

$$\sigma_{\mathcal{L}_\beta}(S_\theta) = \gamma_\theta$$

we have:

$$\mathcal{F}(A) = \sigma^*_{\mathcal{L}_\beta}(\Phi).$$

$\mathcal{F}$ *is* predicatively expressible in $\mathcal{L}_\beta$ *if and only if there is a* $\Phi \in \mathcal{L}_\beta$ *such that* Φ *predicatively expresses* $\mathcal{F}$.

Before proving the main result, the following observation is worth making:

LEMMA 6.3. *Given any set of semantic values* Δ *(finite or infinite) and any total* α*-ary function (*α any *ordinal)* $\mathcal{F}$ *mapping* α*-sequences of members of* Δ *to members of* Δ*,* $\mathcal{F}$ *is connectively expressible in* $\mathcal{L}_\beta$ *if and only if* $\mathcal{F}$ *is predicatively expressible in* $\mathcal{L}_\beta$.

[44] The reader is warned not to confuse this use of "predicative" with the somewhat different, technical use of this term in the literature on second-order logic.

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

Proof. ($\rightarrow$) Assume Φ connectively expresses $\mathcal{F}$ in $\mathcal{L}_\beta$. Let Φ_2 be the result of uniformly replacing all occurrences of $T(S_\gamma)$ (γ any ordinal) with P_1. That is:

$$\Phi_2 = \Phi[T(S_\gamma)/P_1]$$

Φ_2 connectively expresses $\mathcal{F}$ (since, by the connective expressivity of Φ, any assignment $\sigma_{\mathcal{L}_\beta}$ (and corresponding valuation) agrees with $\mathcal{F}$ regardless of what it assigns to any S_γ). Let Φ_3 be the result of uniformly replacing each occurrence of P_γ with $T(S_\gamma)$ in Φ_2. That is:

$$\Phi_3 = \Phi_2[P_\gamma/T(S_\gamma)].$$

Then Φ_3 predicatively expresses $\mathcal{F}$ in $\mathcal{L}_\beta$

($\leftarrow$) Assume Φ predicatively expresses $\mathcal{F}$ in $\mathcal{L}_\beta$. Let Φ_2 be the result of uniformly replacing all occurrences of P_γ (γ any ordinal) with $T(S_1)$. That is:

$$\Phi_2 = \Phi[P_\gamma/T(S_1)]$$

Φ_2 predicatively expresses $\mathcal{F}$ (since, by the predicative expressivity of Φ, any assignment $\sigma_{\mathcal{L}_\beta}$ (and corresponding valuation) agrees with $\mathcal{F}$ regardless of what it assigns to any P_γ). Let Φ_3 be the result of uniformly replacing each occurrence of $T(S_\gamma)$ with P_γ in Φ_2. That is:

$$\Phi_3 = \Phi_2[T(S_\gamma)/P_\gamma].$$

Then Φ_3 connectively expresses $\mathcal{F}$ in $\mathcal{L}_\beta$[45] □

As we have already noted, neither the formalism of (Cook 2008) nor the formalism of (Schlenker 2010) is expressively complete in the relevant sense—each system fails to allow for the definition of notions that are taken as primitive in the other system. The system developed here subsumes both earlier systems, in that all notions primitive in those systems are fully equivalent to defined connectives in the present system. This already constitutes a substantial improvement on either of the earlier treatments, but as already advertised, we can do even better than this, via our proof of a very general expressive completeness result. Before doing so, however, it is worth noting that (Schlenker 2010) contains a similar-in-spirit, although much weaker result, which we can restate in the terminology introduced above as follows ($\mathcal{L}_\beta^{Sc}$ will be used to denote the members of the somewhat different—and expressively impoverished, relatively speaking—transfinite hierarchy of languages studied in (Schlenker 2010)):

THEOREM 6.4. *Given any finite set of semantic values Δ and any n-ary total function (n finite) $\mathcal{F}$ mapping n-tuples of members of Δ to members of Δ, there is an $\mathcal{L}_\beta^{Sc}$ such that $\mathcal{F}$ is connectively expressible in $\mathcal{L}_\beta^{Sc}$.*

Proof. See (Schlenker 2010): 391–393. □

[45] Note that two applications of Lemma 6.3 (one in each direction) allow us to obtain *pure* formulas in the following sense:

- A formula Φ is a *pure connective expression* of a function $\mathcal{F}$ if and only if Φ contains no occurrences of $T(S_\beta)$ (for any $\beta \in ON$).
- A formula Φ is a *pure predicative expression* of a function $\mathcal{F}$ if and only if Φ contains no occurrences of P_β (for any $\beta \in ON$).

The recipe for constructing formulas that connectively express functions given in Theorem 6.5 is pure in this sense.

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

With the present system we can do rather better:

THEOREM 6.5. *Given any set of semantic values* Δ *(finite or infinite), and any* α*-ary total function (*α any *ordinal)* $\mathcal{F}$ *mapping* α*-sequences of members of* Δ *to members of* Δ*, there is a* $\mathcal{L}_\beta$ *such that* $\mathcal{F}$ *is connectively expressible in* $\mathcal{L}_\beta$*.*

Proof. Let Δ be a set of truth values of cardinality κ, and $\mathcal{F}$ a total α-ary function mapping sequences of members of Δ to members of Δ. We will work in $\mathcal{L}_\beta$ where:[46]

$$\beta = max\langle \kappa^{|\alpha|}, sup\{\gamma : \rho_\gamma \in \Delta\} + 1\rangle,$$

which guarantees that all resources required in the construction are available. Let $\langle S_1, S_2, \cdots\rangle$ be a sequence of sentence names of length α. First, we construct, for each sequence of truth values $\Sigma = \langle V_1, V_2, \cdots\rangle$ from Δ of length α, a sentence that *characterizes* Σ—that is, a sentence that is true if $\langle S_1, S_2, \cdots\rangle$ are assigned the respective semantic values from Σ, and false otherwise. Given such a sequence of truth values $\Sigma = \langle V_1, V_2, \cdots\rangle$ let:

$$\begin{aligned} \Phi_{V_\gamma} &= P_\gamma && \text{If } V_\gamma = \top. \\ &= \neg P_\gamma && \text{If } V_\gamma = \bot. \\ &= C^\lambda_\beta(P_\gamma) && \text{If } V_\gamma = \rho_\lambda. \end{aligned}$$

Then the following sentence characterizes $\Sigma = \langle V_1, V_2, \cdots\rangle$:

$$Row_\Sigma = \wedge(\{\Phi_{V_\gamma} : V_\gamma \in \Sigma\}).$$

We now construct, for each sequence of semantic values $\Sigma = \langle V_1, V_2, \cdots\rangle$ from Δ of length α, a sentence that *agrees with* $\mathcal{F}$ on Σ—that is, a sentence that receives $\mathcal{F}(\Sigma)$ as semantic value if $\langle S_1, S_2, \cdots\rangle$ are assigned the respective semantic values from Σ, and false otherwise. Given any sequence of truth values $\Sigma = \langle V_1, V_2, \cdots\rangle$ from Δ of length α, let:

$$\begin{aligned} \Psi_\Sigma &= Row_\Sigma && \text{If } \mathcal{F}(\Sigma) = \top. \\ &= Row_\Sigma \wedge C_0 \ (= Row_\Sigma \wedge \bot) && \text{If } \mathcal{F}(\Sigma) = \bot. \\ &= Row_\Sigma \wedge C_\gamma && \text{If } \mathcal{F}(\Sigma) = \rho_\gamma. \end{aligned}$$

For each sequence of semantic values $\Sigma = \langle V_1, V_2, \cdots\rangle$ from Δ, Ψ_Σ agrees with $\mathcal{F}$ on Σ. As a result:

$$\Xi_\mathcal{F} = \vee(\{\Psi_\Sigma : \Sigma \text{ a sequence of length } \alpha \text{ from } \Delta\})$$

connectively expresses $\mathcal{F}$ since, for any sequence of semantic values $\Sigma = \langle V_1, V_2, \cdots\rangle$ from Δ of length α, all disjuncts but one of $\Xi_\mathcal{F}$ will be false, and the remaining one will take the semantic value assigned to Σ by $\mathcal{F}$. □

COROLLARY 6.6. *Given any set of semantic values* Δ *(finite or infinite), and any* α*-ary total function (*α any *ordinal)* $\mathcal{F}$ *mapping* α*-sequences of members of* Δ *to members of* Δ*, there is a* $\mathcal{L}_\beta$ *such that* $\mathcal{F}$ *is predicatively expressible in* $\mathcal{L}_\beta$*.*

Proof. Combine Lemma 6.3 and Theorem 6.5. □

[46] The first constraint is more generous than we actually need. In fact, we only need β to be sufficiently large such that $\aleph_\beta \geq \kappa^{|\alpha|}$. Nothing hinges on this, however.

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

This completes the technical work of the paper. There are, however, some observations worth making. First, the proof of expressive completeness presented in this section appears rather easy and straightforward—a rather unimaginative infinitary generalization of standard proofs of expressive completeness for subsets of connectives in classical (finitary) propositional logic. This, of course, is by design, but it is also somewhat misleading: the real work of the proof occurs in the previous two sections, where we isolated a collection of primitive connectives that are:

1. Jointly strong enough to allow us to express every function.
2. Individually constrained enough (i.e., α-monotonic, for appropriate α) to allow for the construction of models

and we then explicitly constructed a number of defined notions (in particular, the C_γ^β operators) that are required for the proof.

The second thing to note is that the recipe used in the proof is not unique: given an arbitrary α-ary function on a set-sized collection of semantic values Δ, there is more than one way to construct a complex connective (or complex predicate) that expresses that function. These distinct connectives (or predicates) need not be logically equivalent, since they can disagree on some or all of the 'proper-class-many' semantic values outside of Δ.

Finally, the reader might be forgiven for thinking that the result above is overkill. After all, presumably our natural languages (or any possible extension of them) fail to be nearly as expressively powerful as the full sequence of languages $\mathcal{L}_\alpha$ developed above. This is, no doubt, correct (in virtue of purely physical constraints on our language use, if nothing else). But the formal languages and semantics given above, and the expressive completeness result we have made so much of here, are not meant to provide accurate models of the actual expressive resources of the natural languages that we do, ought to, or even could, read or speak. Rather, the point is this: any natural language that we do read or speak can be mapped onto an appropriate (bounded) fragment of the sequence of $\mathcal{L}_\alpha$s, and, as a result, any expressions from natural language can be given a coherent, revenge-free interpretation in terms of that fragment. In short, the point is not to describe accurately all and only those natural languages we can or do use, but rather to provide a single and singular framework rich enough such that any such natural languages we might use can be consistently, revenge-freely represented within the framework.

§7. Final Embrace. We have completed the task set forth at the beginning of the paper— to prove a powerful expressive completeness result within the Embracing Revenge framework in order to demonstrate that the view is able to live up to the methodological maxim that underlies and motivates the view:

> If Φ seems to be coherently expressible in our natural language(s), then, if at all possible, our formal accounts of those languages (and, in particular, our formal theories of truth and related semantic notions) should allow us to express Φ.

The result of the previous section can be rephrased loosely in terms that fit well with the philosophical discussion with which we began the present project: If we require that the Embracing revenge view should be able to express any apparently coherent notion whose expression is not blocked via indefinite extensibility considerations, and we equate indefinite extensibility with being proper class sized (or involving proper class sized notions,

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

collections, or constructions in an essential manner),[47] then Theorem 6.5 amounts to saying that any function from sequences of semantic values to individual semantic values that *should* be expressible according to the Embracing Revenge view (i.e., that involves a *set* of semantic values, and only *set-sized* sequences as arguments) is, in fact, actually expressible on the view. Or, paraphrasing comments found throughout (Cook 2008) and (Cook 2009): Although there is no single language (better: stage in the development of the single indefinitely extendible language) in which we can say everything, the expressive completeness result shows that, for anything we can say, there is some language (better: stage . . .) in which we can say it.

A final note: Given the inexpressibility of concepts like LANGUAGE, TRUTH VALUE, and STATEMENT, one can and should ask which language in the hierarchy is used when proving the results given above. The answer is simple, if somewhat unexpected: $\mathcal{L}_0$. In more detail, all of the theorems proved in the previous sections should be understood as applying directly to the (classical) set-theoretical *models* of the hierarchy of languages constructed in whatever set theory is contained in $\mathcal{L}_0$, and not *directly* to the indefinitely extensible sequence of languages being modeled by such set-theoretic constructions. Of course, the point of all of this theorem proving is that these results about formal models provide insights into the logic of the (indefinitely extensible) hierarchy of natural languages being modeled in an indirect (albeit familiar) sense. But it is important to note that any particular set theory only provides us with (i.e., guarantees the existence of) an initial segment of the entire set-theoretic hierarchy—this is just what it is for the set theoretic hierarchy to be indefinitely accessible. Thus, the theorems given in previous sections only apply, even *indirectly*, to the initial segments of the indefinitely extensible hierarchy of languages that can be modeled in whatever set theory we happen to currently be using (in more detail: to those natural languages corresponding to formal languages $\mathcal{L}_\alpha$ such that the existence of ordinal α is guaranteed by the set theory within which we are working). Putting all of this a bit loosely, we did not, in the previous section, prove a single theorem labelled "6.5", but instead sketched the proofs of an indefinitely extensible collection of such results—one for each of the successively stronger set theories within which we might be working.[48]

Of course, we can successively strengthen our set theory—for example, via the introduction of large cardinal axioms—in order to expand the portion of the hierarchy of natural languages that we are able to model. Nevertheless, we are never able to 'prove' the expressive completeness of the entire hierarchy of natural languages in one go, but this is as it should be. After all, the indefinite extensibility of language that is the core of the view developed here prevents us even formulating the statement of any theorem that genuinely

[47] Note that we are merely accepting the following, independently plausible claim:

> A collection is indefinitely extensible if and only if it is the size of (or involves in an essential way) a proper class.

We are not suggesting that this amounts to an *explanation* or *analysis* of indefinite extensibility.

[48] These connections between the indefinite extensibility of the set theoretic hierarchy and the indefinite extensibility of the hierarchy of languages $\mathcal{L}_\alpha$ can be fruitfully compared to a similar connection drawn between set theory—an ontological hierarchy—and type theory or higher-order logic—an ideological hierarchy—drawn in (Linnebo & Rayo 2012). Linnebo and Rayo argue that these two hierarchies are really merely two perspectives on the same phenomenon, although like (Rayo 2007), and unlike the approach, they treat the linguistic hierarchy in question purely classically. See (Cook 2008) for a further discussion.

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press

quantifies over all sentences, or all languages, or all truth values. Thus, the role of the formal results in this paper (and in (Cook 2008) and (Cook 2009)) can be summed up along lines similar to the points made at the beginning of this section: Given any particular formal model of the Embracing Revenge view, and the initial segment of the hierarchy of natural languages modeled by that construction, there is a set theory within which we can prove an expressive completeness result of the appropriate sort (i.e., an instance of Theorem 6.5). There is no set theory, and no language, within which we can prove an expressive completeness proof for the entire indefinitely extensible hierarchy of languages, however.[49]

BIBLIOGRAPHY

Beall, J. (2006). True, false, and paranormal. *Analysis*, **66**(2), 2–14.
Beall, J. (2008). *Revenge of the Liar: New Essays on the Paradox*. Oxford: Oxford University Press.
Cook, R. (2008). *Embracing Revenge: On the Indefinite Extensibility of Language*, in (Beall 2008), pp. 31–52.
Cook, R. (2009). What is a truth value, and how many are there? *Studia Logica*, **92**, 183–201.
Cook, R. (2013). *Key Concepts in Philosophy: Paradoxes*. Cambridge, UK: Polity Press.
Dummett, M. (1993). *The Seas of Language*. Oxford: Oxford University Press.
Field, H. (2008). *Saving Truth from Paradox*. Oxford: Oxford University Press.
Kripke, S. (1975). Outline of a theory of truth. *Journal of Philosophy*, **72**, 690–716.
Linnebo, Ø., & Rayo, A. (2012). Heirarchies ontological and idealogical. *Mind*, **121**, 269–308.
Priest, G. (1990). Boolean negation and all that. *Journal of Philosophical Logic*, **19**(2), 201–215.
Rayo, A. (2007). *Beyond Plurals*, in (Rayo & Uzquiano 2007), pp. 220–254.
Rayo, A., & Uzquiano, G. (2007). *Absolute Generality*. Oxford: Oxford University Press.
Rosenblatt, L., & Szmuc, D. (2014). On pathological truths, this Review, **7**(4), 601–617.
Russell, B. (1906). *On Some Difficulties in the Theory of Transfinite Numbers and Order Types*, in (Russell 1973), pp. 135–164.
Russell, B. (1973). *Bertrand Russell, Essays in Analysis*. New York: Braziller.
Schlenker, P. (2010). Super-liars, this Review, **3**(3), 374–414.
Tarski, A. (1933). *The Concept of Truth in the Languages of the Deductive Sciences*, in (Tarski 1983), pp. 252–278.
Tarski, A. (1983). In J. Corcoran, editor. *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*. Indianapolis: Hackett Publishing.
Williamson, T. (1998). Indefinite extensibility. *Grazer Philosophische Studien*, **55**, 1–24.

DEPARTMENT OF PHILOSOPHY
UNIVERSITY OF MINNESOTA – TWIN CITIES
E-mail: cookx432@umn.edu

DEPARTMENT OF PHILOSOPHY
RUTGERS UNIVER
E-mail: tourville.nicholas@gmail.com

[49] An early version of this paper was presented at the Oxford-Buenos Aires Workshop on Truth at *Sociedad Argentina de Análisis Filósofico* (SADAF) in Argentina. Thanks are also due to two very helpful referees. Research on the project was supported by a University of Minnesota Undergraduate Research Opportunities Program (UROP) grant during the Spring of 2011.

https://doi.org/10.1017/S175502031600006X Published online by Cambridge University Press